

Escuela Politécnica Superior

20
21

Trabajo fin de grado

Análisis diferenciador entre Alzheimer y Demencia mediante Redes Neuronales Recurrentes



Emilio Cabañero Fernández

Escuela Politécnica Superior
Universidad Autónoma de Madrid
C/ Francisco Tomás y Valiente nº 11

UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR



Grado en Ingeniería de Telecomunicaciones

TRABAJO FIN DE GRADO

**Análisis diferenciador entre Alzheimer y
Demencia mediante Redes Neuronales
Recurrentes**

Autor: Emilio Cabañero Fernández
Tutor: Álvaro García Martín
Ponente: José María Martínez Sánchez

junio 2021

Análisis diferenciador entre Alzheimer y Demencia mediante Redes Neuronales Recurrentes

Autor: Emilio Cabañero Fernández

Tutor: Álvaro García Martín

Ponente: José María Martínez Sánchez



Video Processing and Understanding Lab

Dpto. 111

Escuela Politécnica Superior

Universidad Autónoma de Madrid

junio 2021

Trabajo parcialmente financiado por el Gobierno de España bajo el proyecto
TEC2017-88169-R (MobiNetVideo)



Todos los derechos reservados.

Queda prohibida, salvo excepción prevista en la Ley, cualquier forma de reproducción, distribución, comunicación pública y transformación de esta obra sin contar con la autorización de los titulares de la propiedad intelectual. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (*arts. 270 y sgts. del Código Penal*).

DERECHOS RESERVADOS

© 20 de Febrero de 2021 por UNIVERSIDAD AUTÓNOMA DE MADRID,
Francisco Tomás y Valiente, nº 1,
Madrid, 28049,
Spain.

Emilio Cabañero Fernández

Análisis diferenciador entre Alzheimer y Demencia mediante Redes Neuronales Recurrentes

Emilio Cabañero Fernández

C\Francisco Tomás y Valiente, nº 1

IMPRESO EN ESPAÑA - PRINTED IN SPAIN

A mi familia
A mis amigos
A mis maestros

Un buen libro es una voz
sabia junto a ti.

Abel Pérez Rojas

AGRADECIMIENTOS

En primer lugar, me gustaría dar las gracias a mi tutor, Álvaro García y mi ponente José María Martínez Sánchez, por ser pacientes con mis descuidos y por estar siempre dispuesto a ayudarme con cualquier duda que tuviera. Gracias por permitirme proponer este tema y haber realizado una investigación tan extensa. También querría agradecer tanto a Álvaro, como a todo el equipo del VPU por su ayuda ante las complicaciones que han ido surgiendo.

A mis compañeros de carrera gracias por hacer más amenas las clases y las prácticas, aunque no siempre aprovechásemos el tiempo al máximo. Ha sido una suerte poder contar con ellos durante todo mi paso por la universidad. Hoy en día a pesar de la situación mundial que hemos vivido se han convertido en un pilar fundamental de todo lo que me rodea, han pasado de ser compañeros a ser verdaderos amigos.

Agradecer a mi familia, su apoyo en los momentos más difíciles y la confianza depositada para poder llevar a cabo este proyecto de vida. Si todo se derrumba a tu paso, siempre hay un punto al que volver a comenzar de nuevo.

A todas aquellas personas que me he cruzado en el camino, los que confiaron en mí y los que no, gracias por formar parte de este proceso. Gracias a ellos hoy soy mejor de como empecé este camino. Gracias a todas las personas que han vuelto con el tiempo a enseñarme que se puede aprender de los errores y que el tiempo sabe poner las cosas en su sitio.

Ha sido una carrera de fondo. Y lo mejor está por venir.

Emilio Cabañero Fernández

Junio 2021

RESUMEN

En los últimos años, el uso de la inteligencia artificial está encontrando una gran cabida dentro del sector de la medicina. En particular, en las enfermedades más importantes del siglo XXI como es el Alzheimer. Una enfermedad difícil de detectar ya que puede ser confundida con otras como demencia senil o Parkinson y cuya prevención es fundamental para evitar sus consecuencias.

Esto ha propiciado que un gran número de entidades públicas y privadas, financien proyectos de investigación para poder efectuar una detección temprana. España es uno de los principales países afectados por esta enfermedad y por ello se han desarrollado planes integrales de prevención del Alzheimer, junto con estudios intensivos en la población española sobre la distribución geográfica de la enfermedad y las medidas para manejarla.

En consecuencia en este trabajo de fin de grado hemos decidido centrar nuestra atención en la prevención del Alzheimer y otras demencias mediante el uso de redes neuronales recurrentes a través del *challenge* lanzado en 2017 por la *European Progression Of Neurological Disease Initiative*. En concreto, la actualización y desarrollo a partir de la información proporcionada por el grupo de trabajo *Computational Brain Imaging Group* en el *Singapore National Research Foundation*, bajo la licencia del MIT.

Además hemos proporcionado una visión actualizada del funcionamiento de la infraestructura y el código desarrollado permitiéndolo comparar con los valores obtenidos anteriormente. Esto es gracias a que el *challenge* sigue activo y día a día surgen nuevas implementaciones para abordar un problema tan grande y que no tiene una única solución válida.

Por último, hemos comentado las múltiples posibilidades a continuar tras este desarrollo y algunas líneas de trabajo futuro que podrían permitir un mejor desarrollo y resolución de los problemas planteados. Queremos recordar y enfatizar que es un proyecto de investigación donde gran parte del mismo consiste en probar soluciones que no siempre darán resultados positivos pero que permiten avanzar en la investigación de un problema de gran envergadura.

PALABRAS CLAVE

Alzheimer, *Machine Learning*, Redes Neuronales Recurrentes, demencia senil, deterioro cognitivo

ABSTRACT

In recent years, the use of artificial intelligence is finding a great place in the medicine field. In particular in the most important diseases of the 21st century, such as Alzheimer's. A disease difficult to detect due to it can be confused with others such as senile dementia or Parkinson's and whose prevention is essential to avoid its consequences.

This has led to a large number of public and private entities funding research projects for early detection. Spain is one of the main countries affected by this disease and therefore comprehensive Alzheimer's prevention plans have been developed, together with intensive studies in the Spanish population on the geographical distribution of the disease and measures to manage it.

That is why in this end-of-degree project we have decided to focus our attention on the prevention of Alzheimer's and other dementias using recurrent neural networks through the challenge launched in 2017 by the European Progression Of Neurological Disease Initiative. Specifically, the update and development from the information provided by the Computational Brain Imaging Group working group at the Singapore National Research Foundation, under the license of MIT.

We have also provided an updated view of the operation of the infrastructure and the code developed allowing it to be compared with the values obtained previously. This is due to the challenge still active and new implementations emerge every day to tackle such a big problem that does not have a single valid solution.

Finally, we have commented on the multiple possibilities to continue after this development and some future lines of work that could allow a better development and resolution of the problems raised. We want to remember and emphasize that it is a research project where a large part of the project consists of testing solutions that will not always give positive results but that allow progress in the investigation of such a huge problem.

KEYWORDS

Alzheimer, Machine Learning, Recurrent Neural Networks, senile dementia, cognitive impairment

ÍNDICE

1	Introducción	1
1.1	Motivación	1
1.2	Objetivos	2
1.3	Organización de la Memoria	2
2	Estado del Arte	3
2.1	Alzheimer (EA)	4
2.1.1	Diferencia del Alzheimer frente a otras demencias	4
2.1.2	Biomarcadores	5
2.2	Historia ADNI	5
2.3	TADPOLE <i>Challenge</i>	7
2.4	<i>Deep Learning</i>	8
2.5	Algoritmos de Predicción	10
2.5.1	Predicción constante	11
2.5.2	RNN	11
2.5.3	SVM/SVR	13
2.5.4	LSS	15
2.5.5	LSTM	15
2.6	PyTorch	17
3	Diseño y Desarrollo	19
3.1	Introducción	19
3.2	Adaptación del <i>dataset</i>	19
3.2.1	Explicación de los valores del <i>dataset</i>	21
3.2.2	Actualización de Tadpole Challenge	23
3.3	Desarrollo red neuronal	26
3.3.1	Diseño de la red	26
3.3.2	Rellenado de datos	27
4	Integración, pruebas y resultados	29
4.1	Protocolo experimental	29
4.1.1	<i>Features</i>	30
4.1.2	Evaluación del rendimiento	30
4.2	Comparación entre algoritmos	33

4.2.1	Primer experimento: base de datos original	33
4.2.2	Segundo experimento: base de datos actualizada	34
4.2.3	Tercer experimento: entrenamiento con base de datos actualizada y evaluada con base de datos original	35
4.3	Resolución	37
5	Conclusiones y trabajo futuro	39
5.1	Conclusiones	39
5.2	Trabajo futuro	40
	Glosario, acrónimos y definiciones	41
	Bibliografía	45
	Apéndices	47
A	Fichero <i>features</i> y utilidades de la red neuronal	49
A.1	Explicación desarrollada de las <i>features</i>	49
A.2	Ficheros modificados en el TADPOLE <i>Challenge</i>	50
A.3	Columnas más importantes en TADPOLE <i>Challenge</i>	52

LISTAS

Lista de códigos

3.1	Modificación de fichero "TADPOLE_D1_D2.csv".....	24
-----	--	----

Lista de ecuaciones

2.1	Algoritmo RNN	11
2.2a	Algoritmo <i>Minimal RNN</i> U_t	12
2.2b	Algoritmo <i>Minimal RNN</i> f_t	12
2.2c	Algoritmo <i>Minimal RNN</i> h_t	12
2.3	Solución al problema SVM	14
2.4	Ecuación solución a la predicción de una SVR	14
2.5a	Vector entrada de datos LSS X_t	15
2.5b	Capa oculta LSS h_t	15
2.5c	Vector predicciones mes siguiente LSS \hat{s}_{t+1}	15
2.5d	Variable continua de predicciones mes siguiente LSS \hat{g}_{t+1}	15
2.6a	Ecuación LSTM <i>Forget Gate Layer</i> f_t	16
2.6b	Ecuación LSTM <i>Input Gate Layer</i> i_t	16
2.6c	Ecuación LSTM <i>Cell Gate Layer</i> g_t	16
2.6d	Ecuación LSTM <i>Cell State</i> c_t	16
2.6e	Ecuación LSTM <i>Output State</i> o_t	16
2.6f	Ecuación LSTM <i>Hidden State</i> h_t	16
2.7a	Ecuación LSTM actualizada <i>Forget Gate Layer</i> f_t	17
2.7b	Ecuación LSTM actualizada <i>Input Gate Layer</i> i_t	17
2.7c	Ecuación LSTM actualizada <i>Cell Gate Layer</i> g_t	17
2.7d	Ecuación LSTM actualizada <i>Cell State</i> c_t	17
2.7e	Ecuación LSTM actualizada <i>Output State</i> o_t	17
2.7f	Ecuación LSTM actualizada <i>Hidden State</i> h_t	17
4.1a	Ecuación de área multiclase bajo la curva operativa del receptor (AUC)	31
4.1b	Ecuación mAUC	32
4.2	Ecuación de precisión de clasificación equilibrada (BCA)	32
4.3	Ecuación error medio absoluto (MAE)	32

4.4	Ecuación de puntuación de error ponderada (WES)	32
4.5	Ecuación de precisión de la probabilidad de cobertura (WES)	32

Lista de figuras

2.1	Extracción de características de un escáner MRI mediante Deep Learning	10
2.2	Algoritmos de Predicción	10
2.3	Estructura RNN según la ecuación 2.1	11
2.4	<i>Minimal</i> RNN vs. Tradicional	13
2.5	Ejemplo de SVR para un problema bidimensional, usando un kernel lineal según la ecuación 2.4	14
2.6	Estructura LSS según las ecuaciones 2.5a 2.5b 2.5c 2.5d	15
2.7	Estructura general redes LSTM y GRU	16
3.1	Imágenes del cerebro donde se extraen los valores numéricos de la base de datos . . .	20
3.2	Número de sujetos y <i>timepoints</i> en TADPOLE <i>Challenge</i>	23
3.3	Número de sujetos y <i>timepoints</i> en el <i>dataset</i> actualizado	25
3.4	Estrategias para rellenar los datos que faltan	28
4.1	Rendimiento de los modelos de cada algoritmo promediado en 10 conjuntos de pruebas con la base de datos original	34
4.2	Rendimiento de los modelos de cada algoritmo promediado en 10 conjuntos de pruebas con la base de datos nueva	35
4.3	Rendimiento de los modelos de cada algoritmo promediado en 10 conjuntos de pruebas con la base de datos nueva y evaluado con la base de datos antigua	36

Lista de tablas

3.1	Tabla de datos originales proporcionados por TADPOLE <i>Challenge</i>	21
3.2	Tabla de datos actualizados proporcionados por TADPOLE <i>Challenge</i>	25
A.1	Tabla de pruebas cognitivas de las <i>features</i> usada por la red	49
A.2	Tabla de medidas de la integridad estructural del cerebro de las <i>features</i> usada por la red	50
A.3	Tabla de las medias FDG PET, AV45 PET y AV1451 PET de las <i>features</i> usada por la red	50
A.4	Tabla de valores pertenecientes a los biomarcadores del líquido cefalorraquídeo (CSF) de las <i>features</i> usada por la red	50
A.5	Tabla de ficheros modificados en TADPOLE <i>Challenge</i>	51

INTRODUCCIÓN

1.1. Motivación

El proyecto tiene como principio fundamental el desarrollo de herramientas para la detección y distinción de los diferentes tipos de degeneraciones cognitivas para su posterior tratamiento. Todo ello propiciando a que dicho diagnóstico (*Computer-Aided Diagnosis*, CAD) permita la detección del Alzheimer (EA) en sus primeros estadios, analizando automáticamente imágenes de resonancia magnética (MRI). El enfoque de detección en etapas tempranas posibilitaría el correcto tratamiento del Alzheimer y su distinción con respecto a la demencia senil, lo que permitiría un ahorro de costes económicos gracias a la prevención y al correcto diagnóstico.

A día de hoy no existe un método eficiente para distinguir correctamente el Alzheimer de la demencia senil, ya que el tratamiento de cada uno por separado difiere.

Mediante el estudio de ciertas zonas del córtex temporo-parietal es posible detectar la degradación degenerativa que supone el proceso. De esta forma podremos estipular un proceso de análisis de las imágenes de los pacientes, lo que permitirá el desarrollo y posterior uso como *software* profesional especializado.

Actualmente el Alzheimer lo padecen 47 millones de personas en el mundo y supone un coste global de entorno a 604.000 millones de dólares, sin considerar que más de un 40 % de las personas lo padecen sin ser diagnosticados. Es un problema carente de solución y de gran importancia para la sociedad española tal y como se recoge en el “Plan Integral de Alzheimer y otras Demencias” [9] aprobado en 2019 por parte del Ministerio de Sanidad, Consumo y Bienestar Social, siendo considerada una gran epidemia silenciosa del siglo XXI.

Es por ello, que el desarrollo de herramientas de prevención y distinción es un bien indispensable hoy en día.

1.2. Objetivos

El objetivo fundamental de este trabajo de fin de grado es proporcionar alternativas y herramientas, que permitan despejar la desinformación y desconocimiento que existe en torno a la enfermedad y su diagnóstico. Gracias a la introducción de nuevas tecnologías como el aprendizaje profundo se busca generar un nuevo enfoque con respecto a la enfermedad y sus fases, con la posibilidad de contemplar su desarrollo a la hora de ayudar a especialistas médicos en el análisis y tratamiento de la misma. Mediante el uso de clasificadores y de Redes Neuronales Recurrentes (RNN) [10] se busca ver la evolución de Alzheimer (EA) en pacientes con deterioro cognitivo elevado, sin deterioro cognitivo y con deterioro cognitivo leve.

1.3. Organización de la Memoria

- **Capítulo 1: Introducción.** Motivación, objetivos y organización de la memoria.
- **Capítulo 2: Estado del arte.** Alzheimer, historia ADNI, TADPOLE *Challenge*, *Deep Learning*, algoritmos de predicción y *PyTorch*.
- **Capítulo 3: Diseño y desarrollo.** Introducción, adaptación del *dataset*, desarrollo red neuronal.
- **Capítulo 4: Integración, pruebas y resultados.** Protocolo experimental, comparación entre algoritmos y resolución.
- **Capítulo 5: Conclusiones y trabajo futuro.**
- **Glosario, acrónimos y definiciones.**
- **Bibliografía.**
- **Apéndices.**

ESTADO DEL ARTE

Para poder plantear todo el proceso de estudio sobre la patología y el desarrollo de las herramientas de estudio a utilizar necesitamos conocer en concreto cuáles son las diferencias que existen entre los diferentes tipos de demencia, permitiendo así conocer más a fondo la enfermedad del Alzheimer. Siendo crucial el diagnóstico precoz de la patología.

En los últimos años, el desarrollo computacional, la gestión y el procesado de la información ha visto un crecimiento exponencial permitiendo avanzar en el procesamiento de señales y el aprendizaje máquina. Esto ha dado lugar a la creación de algoritmos que mediante el procesado de muestras de datos son capaces de desarrollar el reconocimiento de patrones que, unido a la estadística nos permite desarrollar soluciones a los problemas planteados o alternativas para los casos que no compartan la relación con el resto de los datos.

La creciente apertura de nuevas tecnologías en el campo de la medicina ha permitido el desarrollo de uso de nuevas herramientas para el tratamiento y diagnóstico de afecciones. Esto vinculado a la mejora de los sistemas de captación de imágenes en alta resolución como resonancias magnéticas MRI (*Magnetic Resonance Imaging*) ó EDG-PET (*Positron Emission Tomography*) ha permitido una mejora en la precisión y detección de las patologías.

Todo ello se ve reflejado en modelos de redes neuronales, conformadas por un conjunto de neuronas artificiales intercomunicadas entre ellas que, a partir de operaciones matemáticas entre las entradas y los pesos o parámetros, dan lugar a una función coste. La mejor implementación de dichas operaciones matemáticas permite optimizar las capacidades de análisis para la tarea objetivo que ha sido desarrollada.

De esta forma, están siendo utilizadas para la clasificación y segmentación de áreas en específico del córtex temporo-parietal y así comparar las distintas áreas del cerebro involucradas en el desarrollo de las distintas afecciones que puedan dar lugar a la enfermedad.

2.1. Alzheimer (EA)

El Alzheimer es una enfermedad de carácter neurodegenerativo primario afectado por un desequilibrio entre la producción y destrucción de la proteína β -amiloide, el cúmulo e inclusión intracelular de proteína TAU hiperfosforilada, el hipermetabolismo de la glucosa y la disfunción mitocondrial, dando lugar a la atrofia cerebral lo que supone una pérdida de neuronas y pérdida de conexiones sinápticas, que provoca un desgaste degenerativo y funcional de manera exponencial hasta la muerte del tejido vivo.

2.1.1. Diferencia del Alzheimer frente a otras demencias

La demencia es el estado permanente de disfunción cognitiva, ejecutiva y emocional, siendo la demencia degenerativa la asociada al envejecimiento y relacionada con defectos cognitivos múltiples como la pérdida de memoria y alteraciones en la función ejecutiva. Por estos motivos es importante no asociar la demencia con una enfermedad específica, sino como un conjunto de síntomas que afectan a la capacidad de una persona para realizar las actividades cotidianas de forma independiente. La enfermedad de Alzheimer (EA) es un tipo de demencia, pero no es la única ya que existen muchos tipos y causas diferentes de demencia.

- **Demencia frontotemporal [11]:** consiste en la atrofia cerebral del hipocampo que se ve reflejada en las neuroimágenes como la atrofia temporal asimétrica a diferencia del Alzheimer (EA). Además tiene una leve acumulación de la proteína β -amiloide
- **Demencia con cuerpos de Lewy [12]:** es un tipo de demencia parkinsoniana que se ve reflejada en las neuroimágenes como una atrofia cerebral generalizada no específica. A diferencia del Alzheimer (EA) preserva el hipocampo. Además la proteína β -amiloide aumenta considerablemente.
- **Demencia vascular [13]:** demencia asociada a una enfermedad cerebrovascular que origina atrofia cortical. Puede venir generada por múltiples infartos y puede actuar junto con el Alzheimer (EA).

En definitiva, todos los profesionales coinciden en el uso de la proteína β -amiloide, líquido cefalorraquídeo y resonancia magnética para obtener información de los ventrículos cerebrales y el hipocampo, como medida básica para detectar y diferenciar los distintos tipos de demencia y, en particular, la enfermedad del Alzheimer (EA), siendo ésta la que más base genética tiene, donde los biomarcadores suponen la principal fuente de estudio tal y como vamos a reflejar en la sección 2.2.

2.1.2. Biomarcadores

En este trabajo final de grado vamos a hacer uso de biomarcadores extraídos de imágenes de escáneres cerebrales. En concreto, imágenes de escáneres *Magnetic Resonance Imaging*, MRI y *Positron Emission Tomography*, PET. Esto se traduce en que en todo momento nuestro *dataset* consistirá en valores numéricos adquiridos de ADNI (expuesto en la sección 2.2).

El término biomarcador [14], o marcador biológico, se refiere a una amplia gama de medidas que captan lo que está sucediendo en una célula u organismo en un momento dado. Los biomarcadores son signos médicos objetivos (en contraposición a los síntomas declarados por el paciente) que pueden medirse de forma precisa y reproducible permitiendo medir la presencia o el progreso de la enfermedad, además de los efectos del tratamiento. Los biomarcadores pueden tener características moleculares, histológicas, radiográficas o fisiológicas.

Los ejemplos de biomarcadores incluyen desde la frecuencia cardíaca y la presión arterial, pasando por los análisis químicos básicos, hasta las pruebas de laboratorio genéticas más complejas de la sangre y otros tejidos. Los biomarcadores son medibles y no definen cómo se siente o funciona una persona. La cuestión clave es determinar la relación entre cualquier biomarcador medible y los criterios de valoración clínicos pertinentes.

2.2. Historia ADNI

A continuación describimos la formación del *dataset* a lo largo del tiempo, permitiendo mejorar los valores de estudio decisivos para la detección y prevención del desarrollo del Alzheimer (EA). Describiremos el proceso de recopilación de la información, las entidades involucradas en su creación y desarrollo, la evolución de biomarcadores escogidos para la detección de la enfermedad y el desarrollo futuro.

ADNI, *Alzheimer's Disease Neuroimaging Initiative*, es un organismo público/privado formado en 2004 y fundado por el Instituto Nacional de Envejecimiento de Estados Unidos (*National Institute on Aging*) e instituciones sin ánimo de lucro. Fue fundada en respuesta al desarrollo emergente de tratamientos modificadores de la enfermedad para el Alzheimer, en contraste con las terapias de “mejora cognitiva” desarrolladas anteriormente, que no estaban diseñadas para modificar la progresión de la enfermedad.

Esto sirve de preámbulo para controlar la progresión de la patología de la enfermedad y demostrar que los tratamientos ralentizaban la progresión de la misma, dejando patente que se requería la medición directa de la patología del Alzheimer (EA). Por lo tanto, existía la necesidad de validar y estandarizar los biomarcadores (en contraste con las evaluaciones clínicas / neuropsicológicas) que detectaban la patología del Alzheimer (EA).

Desde su fundación, ADNI ha incluido pacientes con demencia que se cree que se debe al Alzheimer (EA), pacientes con deterioro cognitivo leve (MCI) y controles cognitivamente normales.

Inicialmente, ADNI incluía mediciones de resonancia magnética (varias secuencias), FDG PET (variante del *Positron Emission Tomography* con un contraste diferente) y líquido cefalorraquídeo (*Cerebral Spinal Fluid*, CSF) de punción lumbar, además de evaluaciones clínicas, pruebas neuropsicológicas, evaluaciones genéticas y bancos de plasma / suero. Posteriormente, se agregaron la PET amiloide, la PET tau y la recolección de tejido cerebral de la autopsia para patología. ADNI está conformado por más de 2000 participantes de los EE. UU y Canadá, con seguimiento anual.

Todos los resultados clínicos, cognitivos y de biomarcadores están disponibles para cualquier científico del mundo que solicite los datos [15].

El principal logro de ADNI ha consistido en la validación del “fenotipado de β -amiloide ($A\beta$)” utilizando PET amiloide y mediciones CSF de $A\beta$. Tal fenotipado ha dado lugar a nuevos criterios de diagnóstico para el Alzheimer que se basan en la evidencia de biomarcadores de patología. Estos han reemplazado los criterios anteriores basados en los signos y síntomas clínicos, que son tanto insensibles como inespecíficos. Los logros futuros de ADNI pasan por la validación como biomarcador útil de la PET tau longitudinal y la utilización de muestras de plasma / suero para validar las mediciones de $A\beta$, tau / tau fosforilada y la cadena ligera de neurofilamentos como biomarcadores.

ADNI 1

La primera fase de ADNI se lanzó en octubre de 2004 y durante seis años el Instituto Nacional sobre el Envejecimiento, 13 compañías farmacéuticas y dos fundaciones. Originalmente fue diseñado para encontrar biomarcadores más sensibles y precisos para la detección temprana y el seguimiento del Alzheimer. El estudio ADNI1 recopiló y analizó miles de escáneres cerebrales, perfiles genéticos y biomarcadores de sangre y líquido cefalorraquídeo. El estudio incluyó a 400 sujetos diagnosticados con deterioro cognitivo leve (MCI), 200 sujetos con Alzheimer temprano y 200 sujetos de control de edad avanzada.

En ADNI1 se utilizaron medidas de imágenes cerebrales que incluyen resonancia magnética estructural y PET (ambas FDG-PET, que mide el metabolismo de la glucosa en el cerebro, y un estudio piloto que utiliza PET amiloide utilizando un compuesto radiactivo, el compuesto B de Pittsburgh, que mide la acumulación de amiloide en el cerebro).

ADNI GO

ADNI se amplió en 2009 por 2 años durante la fase ADNI GO, agregando 200 nuevos participantes a ADNI1 con deterioro cognitivo leve temprano (EMCI). El objetivo de esta fase fue examinar los biomarcadores en una etapa más temprana de la enfermedad.

ADNI 2

En 2011, ADNI2 comenzó evaluando a los participantes de las fases ADNI1/ADNI GO además de los siguientes nuevos grupos de participantes: 150 controles a ancianos, 100 sujetos con deterioro cognitivo leve temprano (EMCI), 150 sujetos con deterioro cognitivo leve tardío (LMCI) y 150 pacientes con Alzheimer leve. También se agregó un nuevo grupo, deterioro cognitivo subjetivo (*Significant Memory Concern*, SMC) en ADNI2 para abordar la brecha entre los controles sanos y deterioro cognitivo leve (MCI). Esto fue un criterio fundamental de inclusión ya que se manifiesta antes del deterioro clínico y podría servir como población objetivo para los ensayos de intervención temprana en la enfermedad de Alzheimer (EA). El estudio incluyó a 107 sujetos SMC. Una adición importante a ADNI2 fue la adición del análisis PET amiloide con Florbetapir, que es un proceso de escaneo mediante el uso de flúor-18 como componente radiactivo, que se aplicó a los sujetos ADNI2 y ADNI GO, para mejorar la calidad y precisión de diagnóstico del Alzheimer (EA).

ADNI 3

ADNI3 comenzó en 2016 con el objetivo ampliado de determinar las relaciones entre las características de los biomarcadores clínicos, cognitivos, por imágenes, genéticos y bioquímicos en todo el espectro del Alzheimer. ADNI3 agrega escáneres cerebrales que detectan enredos de proteína tau (tau PET), un indicador clave de la enfermedad. Esta fase también continúa con el descubrimiento, la optimización, la estandarización y la validación de las medidas de los ensayos clínicos y los biomarcadores utilizados en la investigación del Alzheimer.

En ADNI3 se agregaron cientos de nuevos sujetos con MCI, sujetos con Alzheimer leve y controles de edad avanzada, teniendo prevista su finalización en agosto de 2022.

ADNI 4

Actualmente se está solicitando información a organismos públicos y entidades privadas para la ampliación de la base de datos con nuevos biomarcadores como PET radio farmacológicos que permitan describir los fenotipos de los participantes análisis de plasma y análisis CSF (especialmente para $A\beta$, tau, tau fosforilada, cadena ligera de neurofilamento, así como alfa sinucleína y TDP 43, y otros), y biomarcadores digitales que se pueden utilizar en la clínica y / o en el hogar.

2.3. TADPOLE Challenge

En nuestro caso vamos a centrarnos en el desafío TADPOLE challenge [16] desarrollado por el grupo de trabajo *Progression Of Neurodegenerative Disease group* del *Centre for Medical Image Computing*, de la *University College London*. El motivo de su empleo reside en la utilización de la base de

datos ADNI, para componer el desafío y estimular el desarrollo a futuro del *dataset* proporcionado por ADNI. Buscaremos a partir de las bases planteadas por el grupo, actualizar el *dataset* (sección 3.2), permitiendo dar una nueva visión a los algoritmos desarrollados (sección 2.5).

El desafío TADPOLE consiste en un conjunto de datos pertenecientes a la base de datos ADNI (sección 2.2) donde a partir del desarrollo de ADNI 3 se ha realizado el seguimiento de ciertos sujetos a lo largo del tiempo. A partir de esta información los participantes deben entrenar sus modelos y algoritmos pronosticando tres resultados claves para los participantes de transferencia de ADNI-3: diagnóstico clínico, subdominio cognitivo de la escala de evaluación de la enfermedad de Alzheimer (ADAS-Cog 13) y volumen total de los ventrículos, que luego se comparan con mediciones futuras.

La diferencia principal del desafío estriba en que los datos de prueba no existían en el momento de la predicción (se adquirieron posteriormente) y que se centra en el desafiante problema de la selección de cohortes para ensayos clínicos mediante la identificación de progresores rápidos. La primera fase de presentación de TADPOLE estuvo abierta hasta el 15 de noviembre de 2017; desde entonces se han adquirido datos hasta abril de 2019 de 219 sujetos con 223 visitas clínicas y 150 exploraciones por resonancia magnética (MRI), que se utilizaron para la evaluación de las predicciones de los equipos participantes, donde se presentaron 33 equipos con un total de 92 exposiciones.

2.4. *Deep Learning*

Como hemos explicado anteriormente en la sección 2.1.2 en este trabajo, vamos a centrarnos en el uso de biomarcadores. Por este motivo, vamos a requerir el uso de tecnologías que permitan trabajar con un gran volumen de datos y permitan generar soluciones a largo plazo. Por ello, se hace necesario hacer uso de tecnologías de *Deep Learning*.

El *Deep Learning* consiste en un subcampo del *Machine Learning* que conforma una estructura neuronal artificial similar a la estructura de un cerebro humano. Esto es debido a que los algoritmos están conformados por neuronas que se interconectan unas con otras, donde una gran parte del proceso de abstracción se realiza a través de una gran cantidad de capas ocultas de procesamiento no lineal para extraer características de los datos y transformar los datos en diferentes niveles de abstracción.

A su vez, el *Deep Learning* tiene redes capaces de aprender sin supervisión o con supervisión. El aprendizaje no supervisado utiliza algoritmos de aprendizaje automático para analizar y agrupar conjuntos de datos que no están estructurados o etiquetados. Estos algoritmos descubren patrones ocultos y características en los datos sin necesidad de intervención humana para agruparlos.

Los modelos de aprendizaje no supervisado se utilizan para tres tareas principales:

- El **clustering** es una técnica que permite agrupar datos no etiquetados en función de sus similitudes o diferencias. Esta técnica es útil para la compresión de imágenes.

- La **asociación** es otro tipo de método de aprendizaje no supervisado que utiliza diferentes reglas para encontrar relaciones entre variables en un conjunto de datos determinado.
- La **reducción de la dimensionalidad** es otra técnica de aprendizaje que se utiliza cuando el número de características de un conjunto de datos es demasiado elevado, reduciendo el número de entradas de datos a un tamaño manejable al tiempo que preserva la integridad de los datos. A menudo, esta técnica se utiliza en la etapa de preprocesamiento de datos.

En cambio, el aprendizaje supervisado se define por el uso de conjuntos de datos etiquetados. Estos se utilizan en el entrenamiento de algoritmos para diferentes tareas, como puede ser la clasificación de datos o la predicción de resultados con precisión. Utilizando entradas y salidas etiquetadas, el modelo puede medir su precisión y aprender con el tiempo.

El aprendizaje supervisado puede dividirse en dos tipos de problemas:

- Los problemas de **clasificación** utilizan un algoritmo para asignar con precisión los datos de prueba en categorías específicas. Los clasificadores lineales, las máquinas de vectores de apoyo, los árboles de decisión y los bosques aleatorios son todos tipos comunes de algoritmos de clasificación.
- La **regresión** es otro tipo de método de aprendizaje supervisado que utiliza un algoritmo para comprender la relación entre las variables dependientes e independientes. Los modelos de regresión son útiles para predecir valores numéricos basados en diferentes puntos de datos. Algunos algoritmos de regresión populares son la regresión lineal, la regresión logística y la regresión polinómica.

Es por ello, que basándonos en los datos haremos uso de aprendizaje supervisado, ya que buscamos predecir el grado de enfermedad del paciente en el futuro, respecto a los datos clasificados de los que disponemos actualmente.

Según la ordenación de las diferentes capas tendremos como resultados diferentes algoritmos, dando lugar a diferentes tipos de redes neuronales como redes convolucionales (*Convolutional Neural Network, CNN*), redes neuronales recurrentes (*Recurrent Neural Networks, RNN*) y redes neuronales artificiales (*Artificial Neural Networks, ANN*). En nuestro caso, vamos a hacer uso de redes neuronales recurrentes ya que permiten una mayor eficiencia a la hora de trabajar con datos que evolucionan en el tiempo. En un principio, se planteó el uso de redes convoluciones que permiten extraer mucho mejor las características particulares de una imagen [17], pero impedían una mejor predicción a la hora de trabajar con datos correlados en el tiempo. Por ello, y al ser datos tanto longitudinales como transversales, para generar las predicciones hacemos uso en este caso de Redes Neuronales Recurrentes. Más concretamente, con parámetros numéricos ó *features* (apéndice A.1) como Adas 13 cuya función ha sido fundamental para la predicción del Alzheimer (figura 2.1).

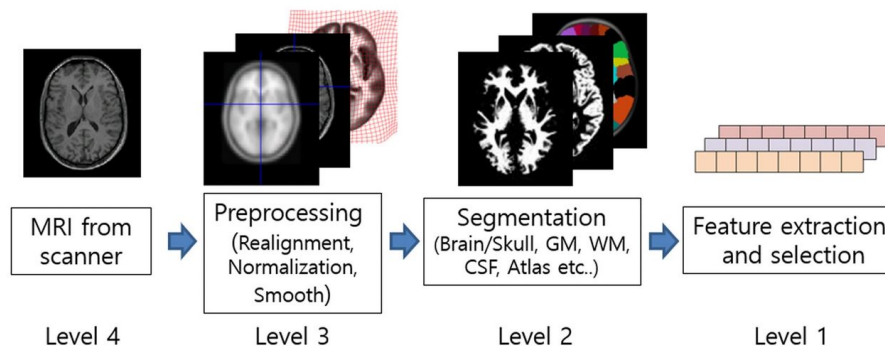


Figura 2.1: Features extraídas de un escáner MRI mediante Deep Learning. Recuperada de [18].

2.5. Algoritmos de Predicción

Una vez expuestas las distintas posibilidades a la hora de escoger una metodología para resolver la problemática planteada, deberemos decidir qué algoritmos queremos emplear para nuestro estudio y comparación de métricas. En razón de la base de datos y el formato de los datos que tenemos vamos a comparar los distintos algoritmos de *Minimal RNN* (*Minimal Recurrent Neural Networks*) [19], una actualización del algoritmo RNN (*Recurrent Neural Networks*) [10], predicción constante, SVM/SVR (*Support Vector Machines/Regression*), LSS (*Linear State Space*) y LSTM (*Long Short-Term Memory*) [20] (figura 2.2).

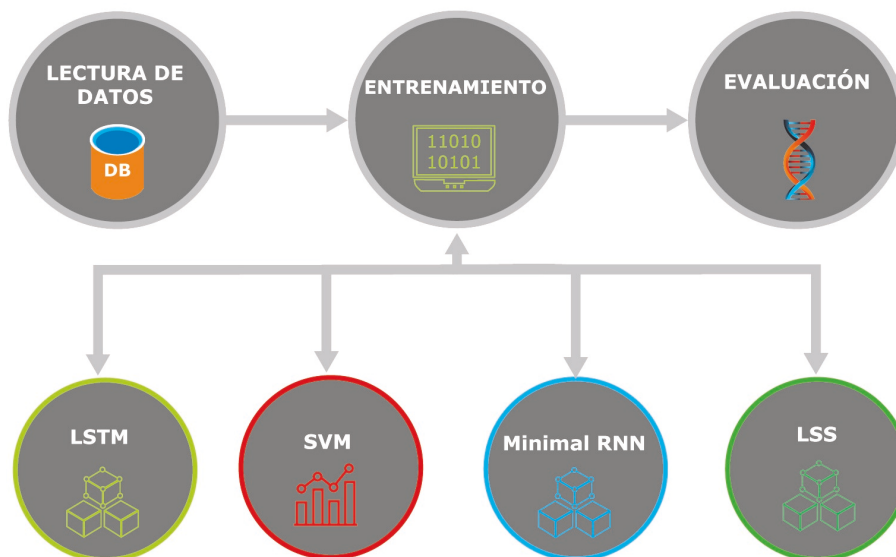


Figura 2.2: Algoritmos de Predicción.

En este trabajo nos vamos a centrar en explicar cuáles son las ventajas de usar *Minimal RNN*, frente a otros algoritmos de predicción más comunes. Por ello, se hará necesario que todos los algoritmos cumplan la misma serie de objetivos y procesos establecidos por *Minimal RNN* para llevar a cabo una comparación justa, como vamos a explicar a continuación algoritmo por algoritmo.

2.5.1. Predicción constante

Un algoritmo de predicción constante permite calcular los valores más próximos sin necesidad de realizar ningún entrenamiento. En nuestro caso, el algoritmo de predicción constante simplemente predice que todos los valores futuros son los mismos que los últimos valores observados. Eso permite que el algoritmo de predicción constante sea bastante competitivo para la predicción a corto plazo y por ello fundamental para efectuar una comparación. Su uso se centrará en utilizarlo como una referencia desde la que partir para llevar a cabo la comparación con el resto de algoritmos.

2.5.2. RNN

Una Red Neuronal Recurrente funciona según el principio de realimentación, es decir, guardar la salida de una capa en particular y conectarla de vuelta a la entrada para así predecir la salida de la capa. Cuando toma una decisión, considera la entrada actual y también lo que ha aprendido de las entradas que recibió anteriormente. Esto hace que a diferencia de las LSTM (sección 2.5.5) sean redes con memoria a corto plazo. Su funcionamiento viene determinado por la ecuación 2.1 y la estructura simple mostrada en la figura 2.3.

$$h(t) = f_c(h(t-1), x(t)) \quad (2.1)$$

donde $h(t)$ consiste en el nuevo estado, f_c es la función con parámetro C , $h(t-1)$ es el estado anterior y $x(t)$ consiste en el vector de entrada dependiente en función del tiempo.

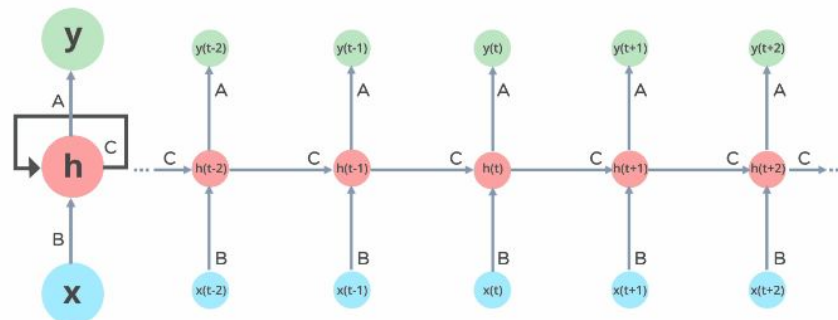


Figura 2.3: Estructura RNN [10] según la ecuación 2.1.

Aquí, x es la capa de entrada, h es la capa oculta e y es la capa de salida. A , B y C son los parámetros de red que se utilizan para mejorar la salida del modelo. En cualquier momento t , la entrada actual es una combinación de la entrada en $x(t)$ y $x(t-1)$. La salida en un momento dado se recupera a la red para mejorar la salida.

Minimal RNN

Observando la figura 2.4, se puede ver la simplificación que supone la implementación del algoritmo *Minimal RNN*. A diferencia de una Red Neuronal Recurrente estándar necesita de un número menor de parámetros y es más eficiente a la hora de predecir con un conjunto de datos menos homogéneo y más reducido de lo habitual. Característica principal por la que ha diferencia de la LSTM es menos propensa a sobreajustar cuando los datos son limitados.

Por lo tanto, *Minimal RNN* es una red neuronal recurrente con una estructura más simplificada y con un desarrollo mejor permitiendo reducir la complejidad y tiempo necesario a la hora de llevar a cabo el entrenamiento. El desarrollo de *Minimal RNN* viene inspirado por el trabajo desarrollado en *Chaos Free Networks* [21] donde se demostraba que pequeños cambios en la entrada de datos podía resultar en cambios dramáticos en el desempeño de la red neuronal dando lugar a que las trayectorias no converjan y generando un sistema mas caótico. Más adelante, deducen que viene producido por la multiplicación matricial que mezcla las dimensiones de un estado.

El funcionamiento de *Minimal RNN* es el siguiente:

- 1.– Escritura para una arquitectura RNN en particular la función actualización de estado.
- 2.– Identifica los términos de la función de actualización que más contribuyen al comportamiento caótico.
- 3.– Elimina el término caótico y lo reemplaza con actualizaciones de estado más simples, una que no implique la mezcla entre diferentes dimensiones del estado.
- 4.– Muestra un rendimiento de precisión similar.

Al reemplazar el estado con actualizaciones más simples, la función de actualización de estado no mezcla ninguna dimensión de su estado de entrada, siendo similar al argumento anterior, y tiene un jacobiano de entrada-salida que tiene valores singulares cercanos a 1 durante el aprendizaje, lo que facilita el entrenamiento. El Jacobiano es una matriz que representa la derivada parcial de una función que se aplica a un vector permitiendo conocer cuánto variará la dimensión de salida con respecto a la dimensión de entrada.

$$u_t = \tanh(W_x \cdot x_t) \quad (2.2a)$$

$$f_t = \sigma(U_h \cdot h_{(t-1)} + W_u \cdot u_t) \quad (2.2b)$$

$$h_t = f_t \cdot h_{(t-1)} + (1 - f_t) \cdot u_t \quad (2.2c)$$

Tal y como podemos ver en la figura 2.4, el codificador $\Phi(\cdot)$ define el espacio oculto, donde la parte recurrente del modelo se limita entonces a moverse dentro de este espacio. En cada paso t , la red

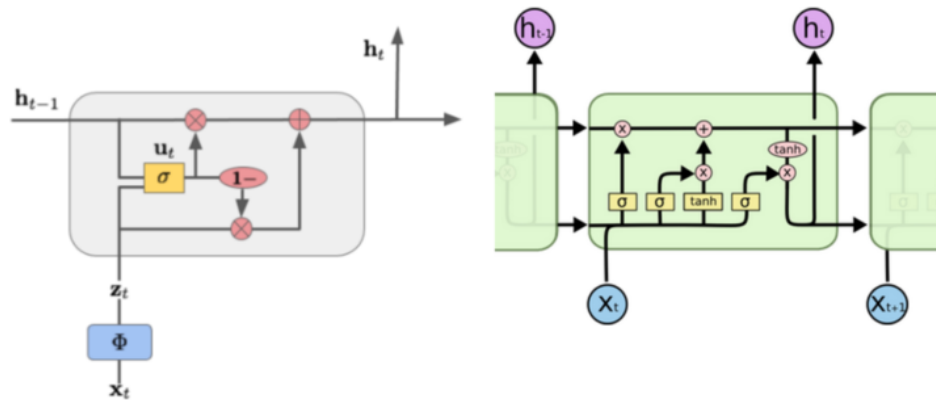


Figura 2.4: *Minimal RNN* vs. Tradicional RNN. Recuperado de [19].

neuronal *Minimal RNN* adquiere su estado anterior $h_{(t-1)}$ y la entrada codificada z_t . El siguiente paso consiste simplemente en generar un promedio ponderado de ambos dependiendo del valor u_t (ó σ en (2.2b)); tal y como podemos observar en (2.2a) y (2.2b). Es decir, la dimensión del estado de salida de la RNN se activa en función de los valores de entrada y se relaja hacia cero si no hay ninguna nueva entrada de esa dimensión. La tasa de relajación viene determinada por el valor σ , que se reactiva una vez se insertan nuevos valores por la entrada codificada z_t .

En comparación con LSTM (*Long Short-Term Memory*), GRU (*Gated Recurrent Unit*), *Minimal RNN* [19] recurre a una actualización mucho más simple dentro de la red neuronal actual, permitiendo conservar el mecanismo de activación, que es fundamental para preservar las dependencias de largo alcance en las RNN y la prohibición de mezcla entre diferentes dimensiones del estado. Todo esto junto con la utilización de un número menor de parámetros hace de *Minimal RNN* un algoritmo mucho más eficiente.

2.5.3. SVM/SVR

La SVM (*Support Vector Machines*) consisten en un conjunto de métodos de aprendizaje supervisado que se utilizan para la clasificación, la regresión y la detección de valores atípicos. Más concretamente en encontrar un hiperplano en un espacio N-dimensional (N corresponde al número de características) que clasifique claramente los diferentes datos obtenidos. Los hiperplanos son límites de decisión que ayudan a clasificar valores obtenidos dando lugar a diferentes clases dependiendo de la dimensión del hiperplano según el número de características.

Los vectores de soporte son puntos de datos que están más cerca del hiperplano e influyen en la posición y orientación del hiperplano. Usando estos vectores de soporte, maximizamos el margen del clasificador y eliminarlos modificará la posición del hiperplano.

Una SVM también se puede utilizar como método de regresión debido a que la SVR (*Support Vector Regression*) utiliza los mismos principios que la SVM para la clasificación, con solo algunas diferencias menores. En primer lugar, debido a que la salida es un número real, resulta muy difícil predecir la información disponible, que tiene infinitas posibilidades. En el caso de la regresión, se establece un margen de tolerancia ε permitiendo considerar los puntos que están dentro del umbral/frontera de decisión y se busca minimizar el error, individualizando el hiperplano que maximiza el margen y teniendo en cuenta que se tolera parte del error.

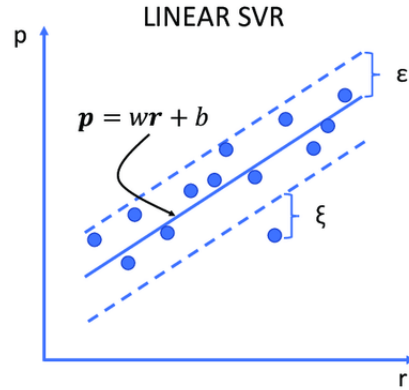


Figura 2.5: Ejemplo de SVR para un problema bidimensional, usando un kernel lineal según la ecuación 2.4. Donde ε_i y ε_i^* corresponden al umbral de tolerancia a ambos lados del kernel. Recuperado de [22].

Dados vectores entrenados SVR resuelve el problema del umbral de decisión del siguiente modo:

$$\frac{1}{2}w^2 + C \sum_{i=1}^N (\varepsilon_i + \varepsilon_i^*) \quad (2.3)$$

Sujeto a las siguientes condiciones siendo $x_i, y_i \in \mathbb{R}$:

$$\left. \begin{array}{l} y_i - wx_i - b \leq \varepsilon_i + \epsilon \\ wx_i + b - y_i \leq \varepsilon_i^* + \epsilon \end{array} \right\} \forall \varepsilon_i, \varepsilon_i^* \geq 0$$

Dando como resultado de la predicción lineal:

$$y = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \cdot K(x_i, x) + b \quad (2.4)$$

2.5.4. LSS

LSS (*Linear State Space*) consiste en un modelo de red proveniente del desarrollo de modelo gráfico probabilístico SSM (*State Space Model*) [23] que describe la dependencia probabilística entre la variable de estado latente y la medición observada. El estado o la medición pueden ser continuos o discretos. Otros términos utilizados para describir los SSM son los modelos ocultos de Markov (*Hidden Markov Models, HMM*) [24] y los modelos de procesos latentes. En nuestro caso, se efectúa una linealización del modelo *minimal RNN* , buscando simplificarlo.

$$x_t = [s_t, g_t] \quad (2.5a)$$

$$h_t = W_h h_{(t-1)} + W_x x_t \quad (2.5b)$$

$$\hat{s}_{t+1} = \text{softmax}(W_s h_t) \quad (2.5c)$$

$$\hat{g}_{t+1} = W_g h_t + g_t \quad (2.5d)$$

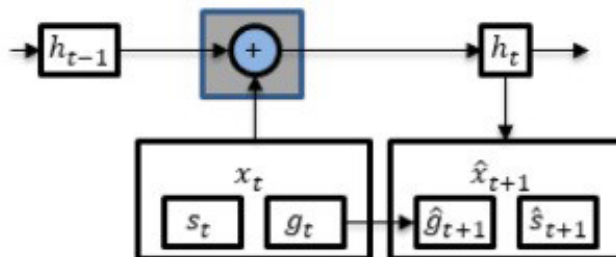


Figura 2.6: Estructura LSS según las ecuaciones 2.5a 2.5b 2.5c 2.5d. Recuperado de [25].

Los parámetros de entrada x_t en la ecuación 2.5a, corresponden con variables categóricas ó de diagnostico (ej, fecha de toma de muestra, edad, sexo, etc.), s_t y variables continuas g_t . Estas a su vez se combinan con el estado oculto anterior $h_{(t-1)}$ (ecuación 2.5b) y dan lugar al modelo predicción del diagnóstico del mes siguiente \hat{s}_{t+1} y variables continuas \hat{g}_{t+1} , utilizando el estado oculto h_t (ecuaciones 2.5c y 2.5d)

2.5.5. LSTM

Una LSTM (*Long Short-Term Memory*), consiste en un tipo especial de RNN capaz de memorizar a corto y largo plazo. Introducidas por Hochreiter y Schmidhuber [20] fueron diseñadas explícitamente para evitar el problema de la dependencia a largo plazo en las RNN, permitiendo preservar la información pasada.

A rasgos generales una LSTM comienza mirando el estado inicial de la capa oculta h_{t_1} y los valores de entrada x_t . La función sigmoidea genera un número entre 0 y 1 para cada número en la celda de estado C_{t_1} (ecuación 2.6a).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.6a)$$

El siguiente paso, es decidir qué información nueva vamos a almacenar en el estado de la celda. Mediante la función sigmoide decidiremos qué valores actualizamos y la capa \tanh genera un vector de nuevos valores candidatos u_t . Después, los combinamos para crear una actualización del estado (ecuaciones 2.6b y 2.6c).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.6b)$$

$$g_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2.6c)$$

Ahora actualizaremos el estado de la celda anterior, c_{t-1} , al nuevo estado de la celda c_t . Para ello multiplicamos el viejo estado por f_t , olvidando los valores que decidimos no seguir almacenando antes al escoger actualizarlos y sumando el vector con nuevos valores candidatos (ecuación 2.6d).

$$c_t = f_t \cdot c_{t-1} + i_t \cdot g_t \quad (2.6d)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2.6e)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (2.6f)$$

Por último, la salida se filtrará mediante el uso de la función sigmoidea decidiendo qué partes del estado vamos a generar. A su vez efectuamos la \tanh al estado de la celda para que los valores estén entre -1 y 1. Estos dos procesos se multiplican entre sí, de modo que solo produzcamos las partes que decidimos (ecuaciones 2.6e y 2.6f).

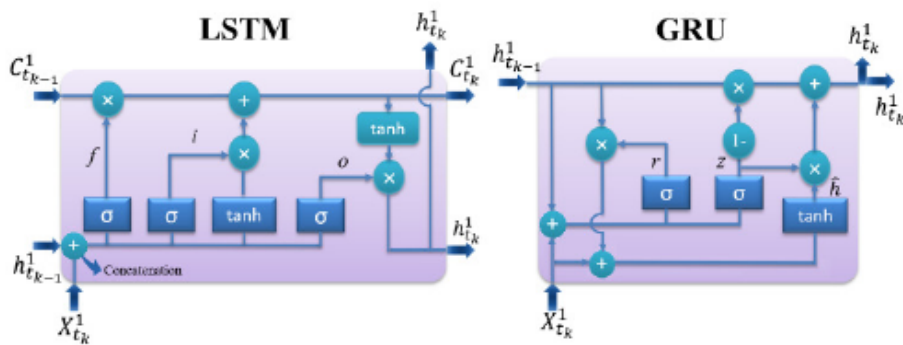


Figura 2.7: Estructura general redes LSTM y GRU. Recuperado de [26].

En el caso del proyecto, se ha buscado dar más énfasis al algoritmo *Minimal RNN* explicado en la sección 2.5.2 debido a que usa menos parámetros que LSTM y, por lo tanto, es menos probable que se sobreajuste. Por ello se ha adaptado el algoritmo LSTM para efectuar una comparación similar al algoritmo *Minimal RNN*.

Como hemos podido comprobar, *Minimal RNN* utiliza menos parámetros que LSTM al eliminar la puerta de salida o_t y establecer que la puerta de entrada o_t sea el complemento de la *forget gate* (es decir, $(1 - f_t)$).

$$f_t = \sigma(U_h^f \cdot h_{t-1} + W_u^f \cdot x_t) \quad (2.7a)$$

$$i_t = \sigma(U_h^i \cdot h_{t-1} + W_u^i \cdot x_t) \quad (2.7b)$$

$$g_t = \tanh(U_h^u \cdot h_{t-1} + W_x^u \cdot x_t) \quad (2.7c)$$

También se elimina la tangente hiperbólica del cálculo de h_t , lo que hace que h_t pase a ser c_t . Además, el término h_{t-1} se elimina del cálculo del término g_t en la RNN mínima, por lo que el estado oculto (h_t) de la RNN mínima decae a cero cuando la entrada (x_t) es cero.

$$c_t = f_t \quad c_{t-1} + i_t \quad g_t \quad (2.7d)$$

$$o_t = \sigma(U_h^o \cdot h_{t-1} + W_u^o \cdot x_t) \quad (2.7e)$$

$$h_t = o_t \quad \tanh(c_t) \quad (2.7f)$$

Es por ello que en el análisis y adaptación que se llevan a cabo, todas las variables (excepto el diagnóstico clínico) fueron normalizadas. De tal forma que la entrada de cero corresponde a la observación del valor medio. En cambio, el estado oculto de la LSTM puede fluctuar incluso cuando la entrada es cero.

2.6. PyTorch

PyTorch [27] es una biblioteca de aprendizaje automático de código abierto que se utiliza para desarrollar y entrenar modelos de aprendizaje profundo basados en redes neuronales que utilizan *Graphics Processing Unit*, GPU y *Central Processing Unit*, CPU. Está desarrollado principalmente por el grupo de investigación de inteligencia artificial de Facebook. PyTorch se puede utilizar tanto con Python como con C ++, aunque Python es el lenguaje más utilizado y donde su implementación es más estable.

A diferencia de la mayoría de los otros marcos de aprendizaje profundo populares como Tensor-

Flow, que utilizan gráficos de cálculo estático, PyTorch utiliza el cálculo dinámico, lo que permite una mayor flexibilidad en la construcción de arquitecturas complejas. Pytorch utiliza conceptos centrales de Python como clases, estructuras y bucles condicionales, que son muy familiares para nuestros ojos y, por lo tanto, mucho más intuitivos de entender. Esto lo hace mucho más simple que otros marcos como TensorFlow que traen su propio estilo de programación.

El empleo de la biblioteca PyTorch ha permitido desarrollar los algoritmos anteriormente descritos en la sección 2.5 permitiendo una mayor rapidez durante la fase de entrenamiento y una mayor eficiencia a la hora de desarrollar las diversas particularizaciones de los algoritmos usando bloques pre diseñados.

DISEÑO Y DESARROLLO

3.1. Introducción

Tras la exposición de los diversos métodos de clasificación y predicción desarrollados en el Estado del Arte, debemos explicar cómo hacemos uso de los datos proporcionados y cuáles son los parámetros más destacables dentro de la base de datos.

Debido a que la información proporcionada por ADNI fue adaptada por TADPOLE *Challenge* 2017 (sección 2.3), nosotros debemos modificar y seleccionar los ficheros necesarios actualizados para poder generar nuestra propia versión renovada de la base de datos que posteriormente deberá ser usada por la red neuronal. Durante este proceso hemos tenido que corregir el código proporcionado por TADPOLE *Challenge*, ya que contenía ciertos errores para poder generar la base de datos previa que se proporcionó en 2017, y que posteriormente hemos actualizado con nuevos datos. La lista de ficheros actualizados se encuentran en el apéndice A.2.

3.2. Adaptación del *dataset*

Como hemos introducido anteriormente el *dataset* viene proporcionado por ADNI y adaptado por TADPOLE *Challenge*. La información usada en un principio por los grupos de trabajo ya viene generada y estructurada en el fichero "*TADPOLE_D1_D2.csv*", pero se proporciona aparte el código implementado y usado para generar el fichero csv. Por este motivo, en primer lugar se han solventado los errores entendiendo el código punto por punto y posteriormente se ha actualizado con nuevos datos.

A continuación, vamos a explicar la estructura de las casillas más significativas de los datos proporcionados para posteriormente explicar las modificaciones necesarias para poder generar el fichero completo y las adaptaciones necesarias dentro de la red para no encontrar dificultades a la hora de entrenarla.

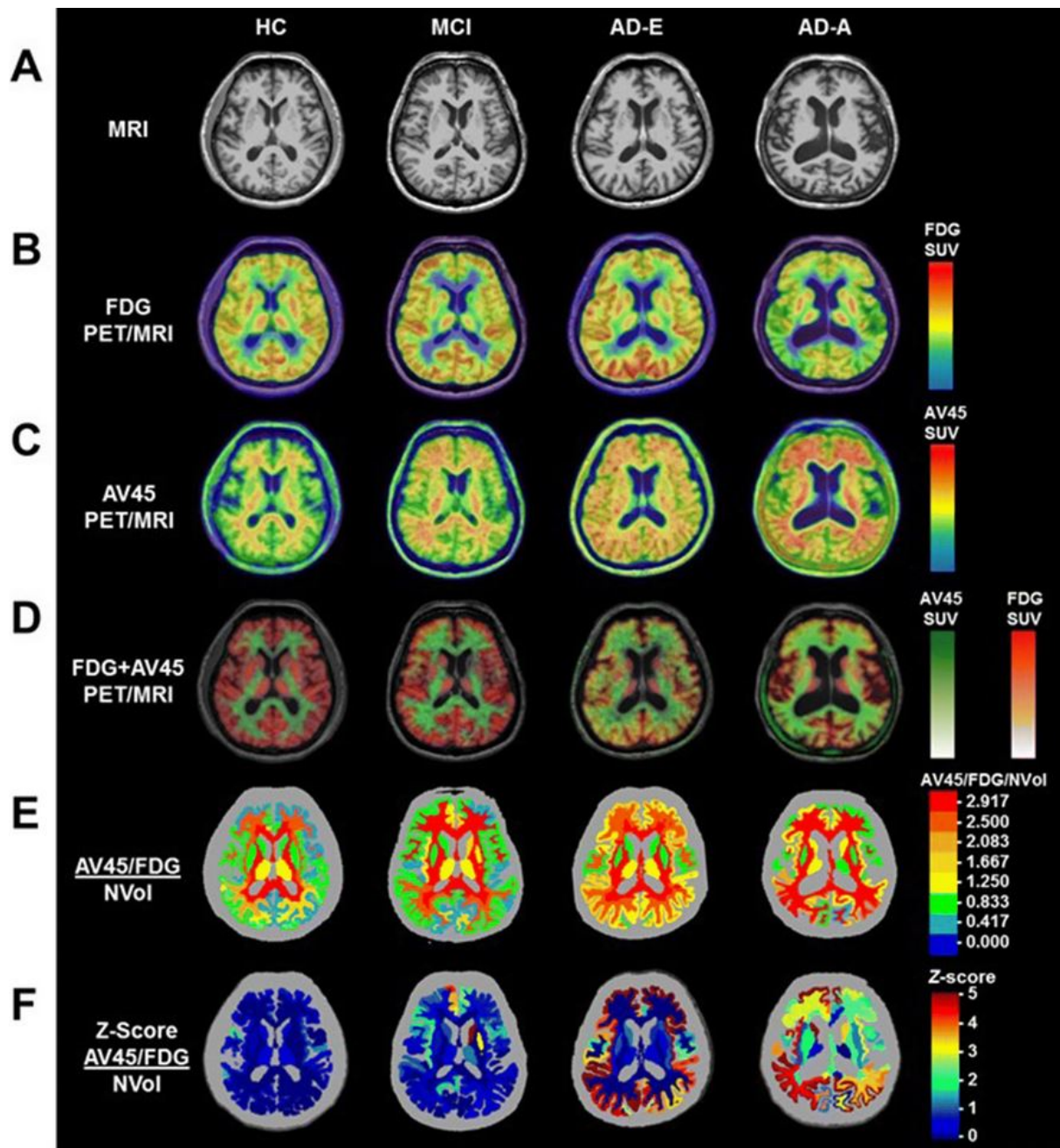


Figura 3.1: Imágenes del cerebro donde se extraen los valores numéricos de la base de datos. Sujetos de control (*Healthy Control*, HC) emparejados con pacientes con trastornos cognitivos leves deterioro (*Mild Cognitive Impairment*, MCI) y pacientes diagnosticados con la enfermedad de Alzheimer en etapa temprana (AD-E) y etapa avanzada (AD-A). Recuperado de [28].

3.2.1. Explicación de los valores del *dataset*

La base de datos consiste en un conjunto de diferentes ficheros proporcionados por los distintos centros hospitalarios que participan en el estudio. En él, se recogen distintos resultados y valores tomados a lo largo del tiempo de diversos pacientes. Algunos pacientes tienen varias entradas a lo largo del tiempo y muchos otros no. Es por ello que se hace crucial el hecho de conseguir predecir tanto longitudinalmente como transversalmente. Todos los ficheros actualizados y el proceso de actualización está reflejado en el apéndice A.2.

Gran parte del conjunto de datos consiste en mediciones tomadas a partir del análisis de imágenes cerebrales y distintos biomarcadores. Imágenes como la resonancia magnética (MRI), la tomografía por emisión de positrones (PET) usando varios biomarcadores como FDG (hipometabolismo), AV45 (amiloide), AV1451 (tau) y así como la imagen por tensor de difusión (DTI). Para las exploraciones de resonancia magnética e imágenes PET se tuvo que promediar y suavizar, incluyendo la corrección de la no uniformidad y el afilado de picos. De este modo se unificaron los datos extrayendo las características regionales más significativas, como el volumen y el grosor cortical, utilizando Freesurfer. Aparte de las mediciones tomadas por imagen se le agregan biomarcadores como el ADAS-Cog 13, proteína determinante para detectar la enfermedad del Alzheimer (EA), información genética e información demográfica general como edad y sexo.

Todo esto en su conjunto constituye a la base de datos como un gran fichero con una cantidad ingente de datos que en muchos casos podrá estar desordenado, mal rellenado o incluso vacío como veremos en la sección 3.2, debido a la gran cantidad de fuentes y a la diferencia de tiempo transcurrido entre las diversas tomas, junto con la variación del número de parámetros necesarios para dar información cada vez más precisa.

Una parte fundamental del *dataset* consiste en la determinación del estado del paciente según evoluciona su caso y la enfermedad. Como en toda investigación necesitamos tener un grupo de control que no padezca la enfermedad a pesar de tener antecedentes genéticos. Cuando el paciente efectúa la primera visita se recoge su diagnóstico en la columna "Dx_bl" con varias etapas de la enfermedad del Alzheimer (EA).

Patología	Num. de participantes	Num. de registros
CN	417	3821
EMCI	310	2319
LMCI	562	4644
SMC	106	389
AD	342	1568

Tabla 3.1: Tabla de datos originales proporcionados por TADPOLE *Challenge*.

- Normales cognitivamente (*Cognitive Normal*, CN)
- Preocupación significativa por la memoria ó fase preclínica del Alzheimer (*Significant Memory Concern*, SMC)
- Deterioro Cognitivo Leve Temprano (*Early Mild Cognitive Impairment*, EMCI)
- Deterioro Cognitivo Leve Tardío (*Late Mild Cognitive Impairment*, LMCI)
- Enfermedad del Alzheimer (*Alzheimer's Disease*, AD)

Un ejemplo claro del aspecto de los distintos tipos de imágenes cerebrales, junto con los diversos grados o estados de la enfermedad del Alzheimer (EA), que posteriormente se traducen en datos numéricos sobre los que trabajamos en el *dataset*, están reflejados en la imagen 3.1. Esta permite ver visualmente la atrofia desarrollada por los pacientes y la importancia de su uso cuantitativo dentro de la base de datos.

Posteriormente, para simplificar el proceso de predicción se simplifican las diversas etapas en tres grupos mayoritarios, reflejados en la columna "Dx". Además se explica la evolución del paciente determinando si ha pasado de una etapa a la siguiente. Los tres grandes grupos consisten en:

- Normales cognitivamente (*Cognitive Normal*, NL)
- Deterioro Cognitivo Leve (*Mild Cognitive Impairment*, MCI)
- Demencia (*Dementia*)

Otras columnas de suma importancia son la columna "RID", donde se proporciona el número del paciente; la columna "EXAMDATE" indicando la fecha en la que el sujeto visitó la clínica cuando el dato fue tomado y las columnas "ADAS13" y "Ventricles", donde se proporcionan los valores de la puntuación compuesta de la prueba cognitiva y la información ventricular, respectivamente.

El *dataset* antiguo proporcionado por TADPOLE *Challenge* [16] consiste en un fichero obtenido por el procedimiento descrito en el apéndice A.2, "TADPOLE_D1_D2.csv" que contiene 12.741 registros de exámenes de 1.737 participantes (tabla 3.1), de los cuales 1.677 sujetos (figura 3.2) son los utilizados para realizar la predicción longitudinal. Es decir, este fichero consiste en la unión de los diferentes conjuntos de datos, donde D1 contiene datos longitudinales de todo el historial de ADNI. D2 consiste en el conjunto de predicción lineal que contiene todos los datos disponibles de los individuos de ADNI *rollover*, es decir, en revisiones anteriores del diagnóstico. D3 consiste en el conjunto de predicción transversal que contiene un único punto temporal, en este caso el más reciente, y un conjunto limitado de variables de cada individuo en revisiones previas. Por último, D4 contiene visitas de sujetos de ADNI *rollover* después del 1 de enero de 2018, donde se incluyen al menos uno de los siguientes campos (estado de diagnóstico, puntuación ADAS o volumen ventricular de la RMN) debido a que este conjunto de datos no existía en el momento de presentar las previsiones en la primera entrega de TADPOLE *Challenge* [16] en Junio de 2019.

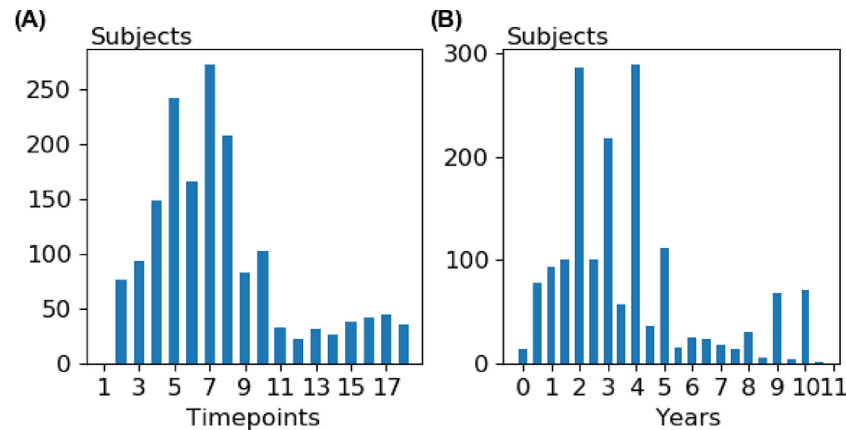


Figura 3.2: (A) Distribución del número de puntos temporales para todos los sujetos del conjunto de datos. (B) Distribución del número de años entre el primer y el último punto temporal para todos los sujetos del conjunto de datos de TADPOLE Challenge. Recuperado de [25].

Toda información extra referente a los parámetros usados por la red neuronal provenientes de las diversas pruebas realizadas a los pacientes quedan reflejadas en la sección 4.1.

3.2.2. Actualización de Tadpole Challenge

Como hemos comentado anteriormente los datos del *dataset* provienen de diversas fuentes y han sido recopilados en diferentes momentos, dando lugar a que ciertos parámetros no sean comunes o no estén formateados del mismo modo. Por lo tanto, han sido modificados con respecto a los proporcionados por ADNI y siguiendo las pautas establecidas por TADPOLE Challenge [16], dando lugar a diversos problemas durante el desarrollo que merecen ser explicados para comprender las soluciones a las que hemos llegado.

En consecuencia, se ha realizado una readaptación de los datos previo al entrenamiento del algoritmo. Primeramente comentaremos las soluciones y pruebas llevadas a cabo en el *dataset* y posteriormente explicaremos en el siguiente apartado 3.3 las modificaciones realizadas en la red neuronal para la correcta aceptación de las modificaciones.

La nueva base de datos en un principio consistía en un *dataset* con una extensión total de 15.102 registros pero como comentaremos a continuación se vería reducida, de tal modo que la nueva base de datos consiste en un *dataset* con una extensión total de 11.377 registros de exámenes de 2.019 participantes (figura 3.3), de los cuales 1.536 sujetos son los utilizados para realizar la predicción longitudinal. El número de sujetos empleados es menor que los proporcionados por el TADPOLE Challenge original, y es debido a que gran parte de los datos han tenido que ser descartados por el motivo explicado a continuación.

Uno de los principales errores a la hora de adaptar la base de datos consistió en la columna

Código 3.1: Código representación de la adaptación mediante pandas de parametros erroneos en fichero "TADPOLE_D1_D2.csv".

```

1  #selecting_characteristics_
2  columns_=['RID','DXCHANGE','EXAMDATE']_
3  data_=pd.read_csv("TADPOLE_D1_D2.csv")_
4  print(data.shape[0])_
5
6  #_making_data_frame_from_csv_file_
7  df_=pd.DataFrame(data)_
8
9  #deleting_values_that_not_fit_with_the_network_
10 indexNames_=df[(df['Month_bl']_==_ ' '),df['Month_bl']_==_'0'),df['Month_bl']_==_'0")].index_
11
12 #reorder_data_by_ID_and_date_
13 df.sort_values(by=['RID','EXAMDATE'],ascending=[True,False],inplace=True)_
14 df.drop(indexNames_,inplace=True)_
15
16 print(df)_
17
18 df.to_csv('output.csv')_

```

“Month_bl”, que como se comenta en el apéndice A.3, consiste en un valor fraccionario de los meses de examen. Debido a que los datos nuevos proporcionados carecían en algunas casillas de valor se optó por rellenarlos con ceros, para solventar los problemas a la hora de entrenar la red. Esta solución no resultó válida y por lo tanto se tuvieron que retirar todas las casillas con valor nulo o vacías, ya que bajo nuestra opinión carecía de sentido mantener columnas en las que el valor de los meses de examen fuese nulo.

De este modo, haciendo uso del código apartado 3.2.2, leemos las columnas que nos proporcionan la identificación del paciente y la fecha de examen, junto con las modificaciones comentadas anteriormente en “Month_bl” donde eliminamos todas las casillas innecesarias para nuestro estudio. Esto da lugar a un fichero de salida ordenado por identificación de paciente y la fecha de examen del paciente de manera ascendente, junto con el resto de casillas que no han sido modificadas.

Posteriormente, se sopesó la posibilidad de retirar las casillas vacías que correspondían a las *features* descritas en el apéndice A.1. En particular, las casillas vacías referentes al hipocampo, el lóbulo temporal y los ventrículos, pero finalmente se descartó ya que se efectuaron varias comprobaciones a nivel de código que demostraron que no era necesario puesto que se descartan al cargarlas.

A raíz de la importancia que tiene la fecha de examen dentro del entrenamiento de la red, se modificó inicialmente el formato en el que se proporcionaba la fecha a la red. Pero seguidamente, se deshizo este cambio debido a que se planteó una mejora en la red neuronal que permitiese leer varios formatos de fechas, sin necesidad de estar manipulando la base de datos más de lo debido y exponiéndose a un mayor número de errores, ya que además se busca la facilidad de implementación

para que futuros grupos de trabajo no encuentren complicaciones ante próximas modificaciones.

Por último, debido a los problemas que presentan los nuevos ficheros proporcionados por ADNI, que incluyen todo tipo de información nueva extra, tal y como se menciona en la sección 2.2, se tuvo que efectuar la modificación de los valores de la columna “Dx” que indican el estado de la enfermedad de los participantes en 3 categorías distintas correspondientes a NL, MCI y *Dementia* (explicado en apéndice A.3). El motivo de su modificación resultó ser que los datos proporcionados por ADNI en primer lugar contenían el campo CN referente a *Cognitive Normal* en vez del campo NL, y cuya modificación por parte del grupo de TADPOLE *Challenge* [16] se efectuó manualmente a posteriori en el fichero *dataset* antiguo. De este modo imitando el proceso llevado por el grupo de TADPOLE efectuamos la modificación de los campos CN a NL.

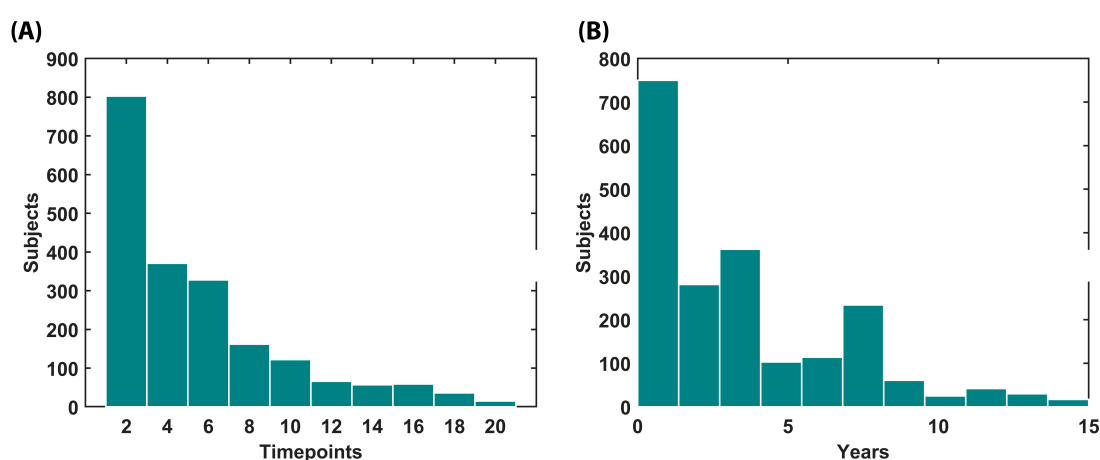


Figura 3.3: (A) Distribución del número de puntos temporales para todos los sujetos del conjunto de datos. (B) Distribución del número de años entre el primer y el último punto temporal para todos los sujetos del conjunto de datos de TADPOLE *Challenge*.

Patología	Num. de participantes	Num. de registros
CN	497	3688
EMCI	330	1883
LMCI	601	4134
SMC	268	646
AD	323	1026

Tabla 3.2: Tabla de datos actualizados proporcionados por TADPOLE *Challenge*.

En nuestro caso, como podemos comprobar en la tabla 3.2 tenemos más valores de sujetos cognitivamente normales y casos intermedios comparado con la tabla 3.1 correspondiente a la base de datos antigua. Esto permite tener un mayor número de casos de estudio, de los sujetos más complicados de predecir, lo que nos proporciona una mayor variedad de casos de estudio a la hora de realizar

el entrenamiento. Respecto al número de registros, ciertamente es menor como se puede observar en la tabla 3.2 que contiene los datos nuevos frente a los datos antiguos representados en la tabla 3.1. La explicación reside en que la nueva base de datos tiene un gran número de entradas nuevas correspondiente con sujetos que solo tienen una única entrada, tal y como se puede observar en la figura 3.3, que representa los datos nuevos gráficamente, frente a la figura 3.2 que representa la base de datos antiguos. Por lo tanto, vinculando la explicación sobre el número de registros, junto con la gran variedad de participantes, obtenemos una base de datos más variada que no particulariza en los extremos sino que trata de equilibrar todo el rango de casos posibles. Además, como explicaremos en la sección 3.3.2 una particularidad fundamental de *minimal* RNN consiste en una novedosa estrategia de relleno del modelo para completar los datos que faltan.

3.3. Desarrollo red neuronal

Tras la exposición del desarrollo de la base de datos, debemos comentar brevemente el funcionamiento de la red neuronal para comprender el funcionamiento y la adaptación llevada a cabo para poder trabajar con la nueva base de datos. En este capítulo además, explicaremos una particularidad fundamental de la red relacionada con la base de datos. Consiste en el rellenado de datos que permite predecir a partir de un punto de tiempo sucesivos puntos correlados.

3.3.1. Diseño de la red

El diseño consiste en cuatro bloques principales, correspondientes con la lectura de la base de datos, el procesado de la información, el entrenamiento del conjunto de algoritmos y la evaluación de los algoritmos.

Las principales modificaciones se efectuaron en el tratamiento y el procesado de la información. En esta parte, tuvimos que procesar y corregir la división de la información referente a los puntos de registro de los pacientes, que posteriormente se usan en el entrenamiento y evaluación (sección 4.1.2). Además, de agregar la posibilidad de que la red procesara fechas en diversos formatos, debido a que el tratamiento de fechas es un aspecto fundamental para el correcto funcionamiento de la red por basar su procesamiento fundamentalmente en puntos de registro y estado de la patología. Posteriormente el sistema clasifica entre los diversos grados de la enfermedad, haciendo uso a su vez de las *features* (sección 4.1.1).

A partir de aquí, se efectúa el entrenamiento de los diversos algoritmos. En este apartado, debemos hacer especial referencia al algoritmo SVM/SVR (sección 2.5.3), encargado de clasificar. Este algoritmo es especialmente dependiente del tamaño del conjunto de datos de entrada. En este caso, se tuvo que adaptar en función de la nueva base de datos, debido a que los límites se establecen de

manera manual y si se sobrepasaran descartaríamos un mayor número de datos de los que han entrado. Además, una característica principal de la red que afecta a la SVM/SVR consiste en la agrupación en periodos de 6 meses debido al formato de recogida de datos por parte de ADNI (sección 2.2). Esta diferencia entre los diferentes puntos de tiempo de un mismo sujeto no son reales y por lo tanto, se adaptó el algoritmo para que los datos intermedios se aproximaran al conjunto espaciado 6 meses más cercano, utilizando de 1 a 4 puntos de tiempo de entrada, para predecir el futuro.

3.3.2. Rellenado de datos

El relleno de datos consiste en una técnica perteneciente al procesamiento de la información. Es especialmente importante ya que como hemos comentado en el apartado anterior la información recogida por ADNI está recogida en un intervalo mínimo de 6 meses, aunque realmente no es así. La falta de información supone un gran problema especialmente para la *minimal* RNN, ya que el TADPOLE Challenge requiere que los participantes hagan una predicción futura en un intervalo mensual y que el rendimiento de la predicción se evalúe con una resolución mensual. Por eso, se plantean las siguientes tres estrategias.

Rellenado hacia adelante

En el relleno hacia adelante se insertan nuevos valores utilizando el último punto de tiempo de los datos disponibles. Por ejemplo, en el 3.4A se pueden observar dos variables de entrada A y B. Los valores de la característica A en el tiempo $t = 2, 3$ y 4 se rellenan utilizando el último valor observado de la característica A (en el tiempo $t = 1$). Lo mismo sucede con los valores en $t = 7, 8$ de la característica A que se rellenan utilizando el valor en $t = 6$, cuando se observa por última vez. En el caso de que falten datos en el primer punto de tiempo, se utiliza el valor medio de todos los puntos de tiempo de todos los sujetos de entrenamiento.

Rellenado lineal

En el caso del relleno lineal se considera que es más eficiente calcular los puntos que faltan realizando una estimación entre los puntos temporales anteriores y los puntos futuros. De este modo, se plantea una interpolación lineal entre el punto de tiempo anterior y el siguiente punto de tiempo a partir de los datos disponibles. Como se puede observar en la figura 3.4B los valores de la característica A en el tiempo $t = 2, 3, 4, 6$ se rellenan utilizando la interpolación lineal. Sin embargo, el relleno lineal no funcionó para los meses 8, 9 y 10 porque no hay datos observados futuros para la interpolación lineal, por lo que se hace uso del relleno hacia adelante para esos puntos de tiempo. Al igual que en el relleno hacia adelante, si faltaban datos en el primer punto de tiempo, se utiliza el valor medio de todos los puntos de tiempo de todos los sujetos de entrenamiento.

Llenado del modelo

En este caso, el planteamiento consiste en utilizar el modelo de la RNN para rellenar los datos que faltan, del mismo modo que se utiliza la predicción de la RNN como entrada para el siguiente punto temporal. En la figura 3.4c se puede comprobar que en el tiempo $t = 2$ a 6, los valores de la característica A se rellenan usando predicciones de la RNN, además de utilizarse para extrapolar características que tienden a finalizar en breve (por ejemplo, el tiempo $t = 8$ y 9).

La ventaja principal del llenado de modelos es que se utilizan los conjuntos completos de características para la predicción del valor nuevo. Por ejemplo, ambas características A y B en el tiempo $t = 1$ se utilizan por la RNN para predecir ambas características de entrada en el tiempo $t = 2$. En cambio, en el relleno directo o lineal, sólo se utiliza una de las características para imputar la contraria. Por ejemplo la característica A para imputar la característica B.

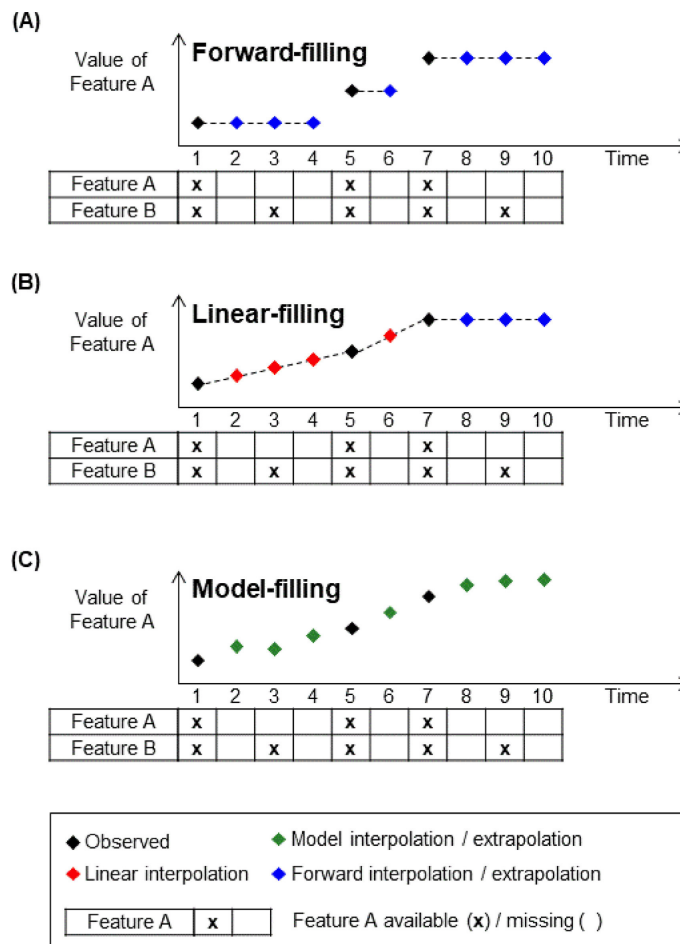


Figura 3.4: Estrategias para rellenar los datos que faltan. **(A)** Rellenado hacia delante de los valores perdidos utilizando el último valor observado. **(B)** Relleno lineal de los valores perdidos utilizando la interpolación lineal entre los valores observados anteriores y los siguientes. Al no existir datos observados futuros para los meses 8, 9 y 10, se utiliza el llenado hacia adelante en estos puntos de tiempo ya que no había datos observados futuros para realizar la interpolación lineal. **(C)** Rellenado de modelo con los valores perdidos utilizando las predicciones del modelo. Recuperado de [25].

INTEGRACIÓN, PRUEBAS Y RESULTADOS

Hemos comentado en el capítulo 3 que la evaluación de resultados viene centrada por cuatro características principales (*Multiclass Area Under the receiver operating Curve*, mAUC; *Balanced Classification Accuracy*, bca; *Mean Absolute Error of Alzheimer's Disease Assessment Scale 13-item Cognitive Subscale*, Adas13 MAE; *Mean Absolute Error of ventricles*, ventricles MAE) que permiten realizar una valoración global de los mejores algoritmos por TADPOLE Challenge (sección 2.3). En este capítulo, se expondrán las cuatro características principales de evaluación y las distintas diferencias que hay entre ellas, además de abordar las diferencias entre los diversos parámetros de entrenamiento y el efecto de estos a la hora de obtener la mejor predicción posible.

Los algoritmos empleados han sido descritos en la sección 2.5 y su entrenamiento se ha realizado según se va a comentar en la sección 4.1.2, llevando a cabo una predicción longitudinal de los sujetos analizados de hasta 6 años en el futuro. Cabe destacar que gran parte de los algoritmos están desarrollados con módulos pertenecientes a la librería *PyTorch*, por lo que el entrenamiento es algo más eficiente gracias a ello. En nuestro caso, el número de *epochs* se ha fijado en 100 al igual que las pruebas efectuadas con la base de datos antigua. Con respecto al *batch size* se ha decidido mantener en 128, el *learning rate* fijarlo en 0.00129066, ya que el rango óptimo se encuentra entre 10^{-5} y 10^{-2} . El hardware utilizado para este fin ha sido un procesador 12 x Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz (2 Sockets), una tarjeta Gráfica NVIDIA TitanX 11Gb y una memoria RAM 80Gb.

4.1. Protocolo experimental

Una vez establecida la modificación realizada en la base de datos y las diferencias implementadas entre la base de datos nueva y antigua en la sección 3.2, debemos comentar el protocolo experimental que vamos a pautar antes de comprobar la eficacia de los diferentes algoritmos en las pruebas y resultados, recogidos en la sección 4.2. Esta sección se ha dividido en dos partes fundamentales, las *features*, donde se recogen variables necesarias para el entrenamiento y testeo de la red, y la evaluación del rendimiento, donde dividimos el conjunto de datos entre entrenamiento, validación y prueba, además de describir las métricas de los valores de evaluación.

4.1.1. *Features*

Las *features* consisten en las características extraídas de la base de datos para realizar la predicción y el entrenamiento de la red. Puede consistir en cualquier característica que sea de valor para determinar dicha predicción, y su elección o descarte puede mejorar o empeorar dicho resultado como observaremos a continuación.

En este trabajo hemos determinado mantener la coherencia utilizando el mismo conjunto de 23 variables usado tanto para entrenar la red con la base de datos antigua en el trabajo realizado por el grupo *Computational Brain Imaging Group* en el *Singapore National Research Foundation* [25] y por el *TADPOLE Challenge* [16]. Las *features* consisten en un conjunto de 23 variables, que incluyen diagnóstico, puntuaciones de pruebas neuropsicológicas, características anatómicas derivadas de la resonancia magnética (MRI) T1, medidas de tomografía por emisión de positrones (PET) y marcadores de líquido cefalorraquídeo (CSF). Para una explicación más detallada acerca de las variables usadas por la red neuronal se recomienda el apéndice A.1 donde se especifica el contenido de las variables usadas.

Como hemos comentado anteriormente la modificación de la elección del número de *features* puede mejorar o empeorar la evaluación de la red y vendrá siempre afectado por la forma en la que queremos abordar nosotros el problema planteado y la solución que damos. Para demostrarlo realizamos una ablación de las características para analizar la contribución de las diferentes características al rendimiento. En concreto, decidimos calcular la contribución del modelo *minimal* RNN entrenado en la predicción a la hora de eliminar una característica de los datos de entrada. El proceso llevado a cabo para realizar la prueba consistió en fijar el valor medio del conjunto de datos al eliminar dicha característica y mantener inalteradas las demás características de entrada. De tal modo que había 23 versiones diferentes de datos de entrada, en las que cada versión tenía una característica diferente eliminada. Por lo tanto, al utilizar el modelo *minimal* RNN entrenado con la división de datos comentada en la sección 4.1.2 se obtuvo una gran caída en el rendimiento de la predicción, dejando patente que cuando se suprime una característica se pierde precisión en la predicción.

En definitiva, todas las características son necesarias para obtener un mayor rendimiento del modelo.

4.1.2. Evaluación del rendimiento

Para evaluar el rendimiento de la red, pautamos que la proporción de sujetos en los conjuntos de entrenamiento, validación y prueba sea de 18:1:1, dividiéndolos de manera aleatoria. De tal forma que para los sujetos de los conjuntos de validación y prueba, la primera mitad de los puntos temporales de cada sujeto se utiliza para predecir la segunda mitad de los puntos temporales del mismo sujeto. Además, para que todas las variables (excepto la categoría de diagnóstico, que era categórica en lugar

de continua) y el conjunto de entrenamiento estén equilibrados se aplica la normalización estadística y así poder hacer comparaciones entre los diferentes elementos. A su vez, del conjunto de entrenamiento se extraen la media y la desviación estándar para normalizar los conjuntos de validación y prueba.

De tal modo, que la división aleatoria de los datos en conjuntos de entrenamiento, validación cruzada y prueba se repite 10 veces garantizando la estabilidad de los datos y teniendo cuidado de que los conjuntos de prueba no se solapen y cubran todo el conjunto de datos. Por ello, la validación cruzada se realiza con el conjunto de datos de entrenamiento, para evitar el sobreajuste de los modelos.

Métodos de evaluación

A la hora de evaluar las predicciones del estado clínico hacemos uso de las siguientes métricas:

- i. El área multiclase bajo la curva operativa del receptor (mAUC).
- ii. La precisión global de clasificación equilibrada (BCA).

En cambio, para la ADAS y el volumen ventricular, utilizamos tres métricas:

- i. Error medio absoluto (MAE).
- ii. Puntuación de error ponderada (WES).
- iii. Precisión de la probabilidad de cobertura (CPA).

El mAUC es independiente del tamaño de los grupos y ofrece una medida global de la capacidad de clasificación que tiene en cuenta las probabilidades relativas asignadas a cada clase. En cambio, el BCA, más sencillo, no aprovecha la naturaleza probabilística de las previsiones, sino que se limita a considerar la precisión de la clasificación más probable. Ambas métricas normalizan el tamaño de los grupos, lo que es importante porque el grupo más pequeño (AD muy probable en el momento de la prueba) es el más importante.

El MAE se centra exclusivamente en la exactitud de la predicción ignorando la confianza, mientras que el WES incorpora las estimaciones de confianza de los participantes en la puntuación del error. El CPA proporciona una evaluación de la precisión de las estimaciones de confianza, independientemente de la precisión de la predicción.

A continuación, realizaremos las formulaciones completas de las predicciones del estado clínico:

1.– Área multiclase bajo la curva operativa del receptor (mAUC):

La clasificación de una clase frente a otra se realiza mediante el AUC:

$$\hat{A}(c_i|c_j) = \frac{s_j - n_i(n_i + 1)/2}{n_i n_j} \quad (4.1a)$$

donde n_i y n_j son el número de puntos que pertenecen a las clases i e j respectivamente,

mientras que s_i es la suma de los rangos de los puntos de prueba de clase i después de clasificar todos los puntos de datos de la clase i y j en probabilidad creciente de pertenecer a la clase i .

$$mAUC = \frac{2}{L(L-1)} \sum_{i=2}^L \sum_{j=1}^i \hat{A}(c_i, c_j) \quad (4.1b)$$

2.– Precisión de clasificación equilibrada (BCA):

$$BCA = \frac{1}{2L} \sum_{i=1}^L \left[\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right] \quad (4.2)$$

donde TP, FP, TN, FN representan el número de verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos para la clasificación como clase i . Los verdaderos positivos son puntos de datos con la etiqueta verdadera i e clasificados correctamente como tal, mientras que los falsos negativos son los puntos de datos con la etiqueta verdadera i clasificados incorrectamente a una clase diferente $j \neq i$. Los verdaderos negativos y los falsos positivos se definen de forma similar. L representa el número de clases.

En cambio las formulaciones de las predicciones de características continuas son:

1.– Error medio absoluto (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N \tilde{M}_i - M_i \quad (4.3)$$

donde N es el número de puntos de datos adquiridos en el momento en que se evalúan las previsiones, M_i es el valor real en el individuo i de los datos futuros, \tilde{M}_i es la mejor suposición del participante sobre M_i .

2.– Puntuación de error ponderada (WES):

$$WES = \frac{\sum_{i=1}^N \tilde{C}_i \tilde{M}_i - M_i}{\sum_{i=1}^N \tilde{C}_i} \quad (4.4)$$

donde las ponderaciones \tilde{C}_i son las confianzas relativas del participante en \tilde{M}_i . Estimamos \tilde{C}_i como la inversa de la anchura del intervalo de confianza del 50 % de su estimación del biomarcador, $\tilde{C}_i = (C_+ - C_-)^{-1}$. $[(C_+, C_-)]$ es el intervalo de confianza del participante.

3.– Precisión de la probabilidad de cobertura (CPA):

$$CPA = |\text{Probabilidad de cobertura real} - \text{Cobertura nominal } p| \quad (4.5)$$

4.2. Comparación entre algoritmos

A continuación, se presentan los resultados de los análisis realizados, donde se han ajustado ciertos hiperparámetros de la configuración de la red con el objetivo de plasmar los resultados que mejor rendimiento tengan. Este procedimiento se ha realizado con todos los análisis llevados a cabo, buscando un punto intermedio para que las comparativas realizadas entre todos los análisis sea equilibrada.

Como hemos comentado al principio de la sección 4 el *batch size* se ha decidido mantener en 128, el *learning rate* fijarlo en 0.00129066, el *input dropout rate* en 0.1, el *recurrent dropout rate* en 0.4, la regularización del peso L2 en 10^{-5} y el número de capas ocultas en 2, junto con un tamaño de capa oculta de 128. Todo esto bajo las recomendaciones del algoritmo de HORD [29], que consiste en un algoritmo que permite encontrar los mejores hiperparámetros maximizando el rendimiento del modelo en el conjunto de validación. Todas estas condiciones se van a mantener a lo largo de la siguientes subsecciones.

Además en aras de la brevedad, denotamos *minimal* RNN como RNN en todas las figuras y tablas posteriores. Para mayor claridad, sólo mostramos *minimal* RNN con relleno de modelo (RNN-MF), LSS con relleno de modelo (LSS-MF), LSTM con relleno de modelo (LSTM-MF) y SVM/SVR utilizando un punto de tiempo de entrada debido a que dan los mejores resultados dentro de las distintas tecnologías de rellenado de datos en los distintos modelos.

4.2.1. Primer experimento: base de datos original

Siguiendo lo comentado al comienzo de la sección 4.2, obtenemos como resultado la figura 4.1 donde viene reflejados los resultados de todos los algoritmos al emplear la base de datos original y se puede observar que el algoritmo que destaca por encima de todos es el de mínima RNN, dando los mejores resultados.

Podemos observar que mínima RNN-MF es estadísticamente mejor que LSS-MF para el diagnóstico clínico, pero no necesariamente mejor para la ADAS-Cog13 o el volumen ventricular. Del mismo modo, mínima RNN-MF es estadísticamente mejor que LSTM-MF para el diagnóstico clínico y el volumen ventricular, pero no para el ADAS-Cog13. Esto se puede deber a que mínima RNN-MF trabaja mejor en predicciones a largo plazo y un mayor número de puntos por lo que se beneficia frente a otros algoritmos como el de predicción constante que trabaja mejor para intervalos cortos. También podemos observar en el diagnóstico clínico que los algoritmos de LSS-MF, LSTM-MF y SVM-MF obtienen valores parecidos, por lo que demuestran predecir bastante bien al tener una base de datos con un gran número de puntos de tiempo por sujeto.

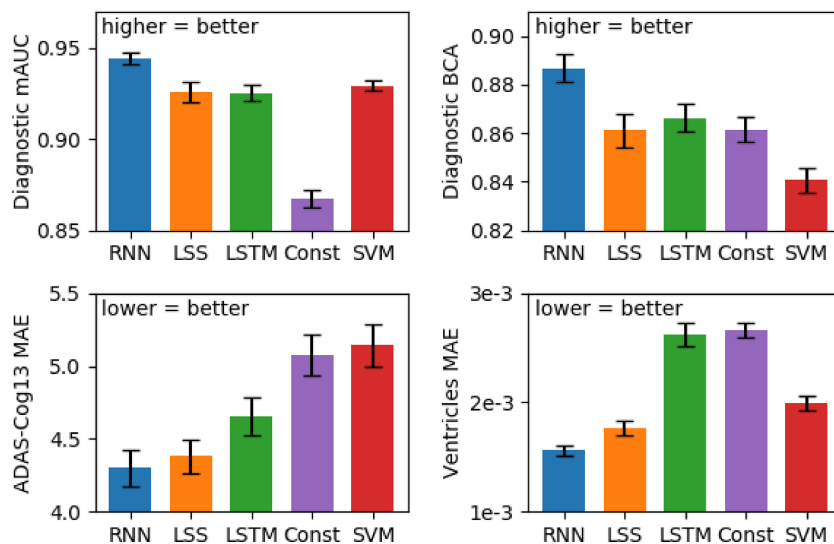


Figura 4.1: Rendimiento de los modelos de cada algoritmo promediado en 10 conjuntos de pruebas con la base de datos original. Las barras de error muestran el error estándar en los conjuntos de pruebas. Para el diagnóstico clínico, los valores más altos de mAUC y BCA indican un mejor rendimiento. Para ADAS-Cog13 y ventrículos, un MAE más bajo indica un mejor rendimiento. En aras de la brevedad, denotamos la RNN mínima como RNN, he indicamos que la SVM/SVR corresponde a 1 punto de tiempo. Recuperado de [25].

4.2.2. Segundo experimento: base de datos actualizada

A partir del modelo definido en la sección 4.2.1, se decide actualizar la base de datos, dando como resultado la figura 4.2. En este caso, tanto la mínima RNN-MF como la SVM/SVM, destacan por tener los mejores resultados que el resto de algoritmos.

Como se puede advertir en la figura 4.1 mínima RNN-MF junto con SVM/SVR son estadísticamente mejores que LSS-MF para el diagnóstico clínico, pero no necesariamente mejor para el volumen ventricular. En cambio, en la ADAS-Cog13 son superiores por una amplia diferencia. Además cabe recalcar que con la base de datos nueva los valores de todos los algoritmos en todas las métricas son más homogéneos, en especial en el volumen ventricular, donde se obtiene una mejora significativa por parte de todos los algoritmos. Esto puede estar mayormente propiciado por el incremento de sujetos con patologías intermedias entre el Alzheimer y la normalidad cognitiva como bien se mostraba en la tabla 3.2 en la sección 3.2.2, junto con la gran cantidad de sujetos nuevos con un solo punto de tiempo tal y como se observaba en la figura 3.3. De este modo, la predicción a largo plazo por parte de mínima RNN-MF es algo más complicada de predecir y por ello, el resultado medio para el diagnóstico clínico en la figura 4.2 es algo menor que el resultado obtenido en la figura 4.1. Si observamos detenidamente la métrica en la ADAS-Cog13 de la figura 4.2 podemos comprobar que ha empeorado notoriamente el resultado obtenido por parte de RNN-MF, LSS-MF y LSTM-MF alcanzando el mismo valor o peor que el de SVM/SVR en la figura 3.2. De este modo, podemos afirmar que la métrica en la ADAS-Cog13 es directamente dependiente del número de puntos de tiempo de un sujeto, y que a mayor número de

puntos mejor será la predicción.

En resumidas cuentas, la actualización de la base de datos deja de manifiesto que mínima RNN-MF está al mismo nivel que la SVM/SVR para el diagnóstico clínico, donde además LSS-MF y LSTM-MF mejoran significativamente gracias al incremento de sujetos con patologías intermedias entre el Alzheimer y la normalidad cognitiva. Pero donde también vemos en la ADAS-Cog13 la dependencia de la cantidad del número de puntos de tiempo en un sujeto.

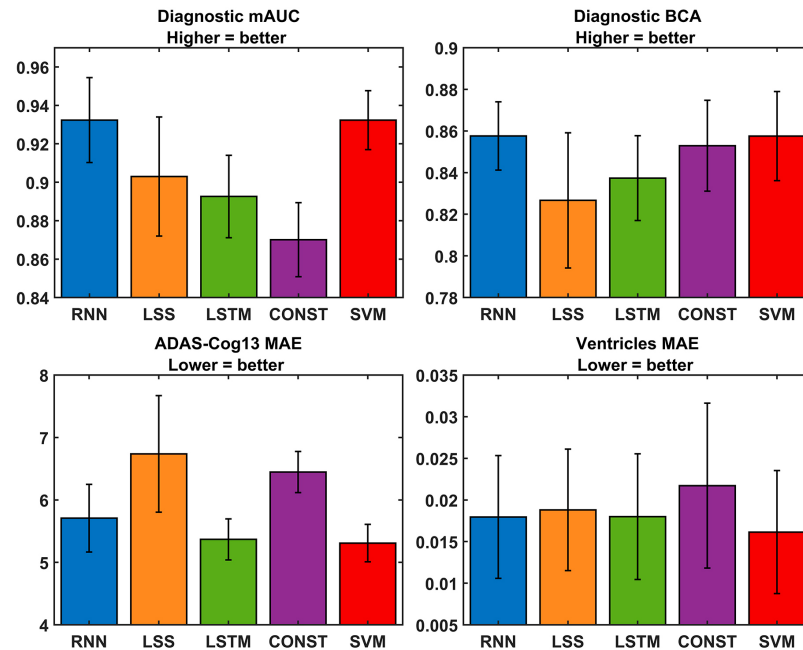


Figura 4.2: Rendimiento de los modelos de cada algoritmo promediado en 10 conjuntos de pruebas con la base de datos nueva. Las barras de error muestran el error estándar en los conjuntos de pruebas. Para el diagnóstico clínico, los valores más altos de mAUC y BCA indican un mejor rendimiento. Para ADAS-Cog13 y ventrículos, un MAE más bajo indica un mejor rendimiento. En aras de la brevedad, denotamos la RNN mínima como RNN, he indicamos que la SVM/SVR corresponde a 1 punto de tiempo.

4.2.3. Tercer experimento: entrenamiento con base de datos actualizada y evaluada con base de datos original

A diferencia de los subapartados anteriores en este se plantea el entrenamiento haciendo uso de la base de datos actualizada y evaluándola con la base de datos original, permitiendo así realizar un análisis cruzado entre las dos bases de datos. A partir del resultado de este análisis obtenemos la figura 4.3.

En primer lugar, podemos observar como todas las métricas han empeorado drásticamente. Esto puede venir influenciado por haber realizado el entrenamiento con una base de datos que tiene un mayor número de puntos temporales correspondiente a la entrada única de varios sujetos individualmente

tal y como se observa en la figura 3.3, frente a la evaluación con una base de datos que tiene un mayor número de puntos de tiempo de un sujeto, además de tenerlos espaciados entre los 2 a 4 años de la primera vez en la que el sujeto comienza a formar parte de la base de datos.

Pese a haber comentado en las secciones 4.2.1 y 4.2.2 que mínima RNN-MF y SVM/SVR tras la actualización de la base de datos eran estadísticamente mejores que el resto de algoritmos, en este caso no se produce. Tal es el caso, que podemos observar que el diagnóstico clínico en la figura 4.3 es algo peor que en la figura 4.2 del apartado anterior, pero no destaca ningún algoritmo por ser especialmente eficiente. Ciertamente la mejora de la LSS-MF, la LSTM-LF y la predicción constante, es bastante significativa en la figura 4.3, cuya explicación puede venir referida a que para los sujetos de prueba, la primera mitad de los puntos temporales de cada sujeto se utilizó para predecir la segunda mitad de los puntos temporales del mismo sujeto tal y como hemos mencionado en la sección 4.1.2. Es decir, un sujeto en la base de datos actualizada tiene un mayor número de puntos temporales y al ser evaluado con la base de datos original que contienen muchos menos puntos temporales de un mismo sujeto, el resultado va a ser mucho mejor. Sorprende el hecho de que a pesar de hacer uso de la técnica de rellenado de valores mediante llenado de modelado en todos los algoritmos, esto no permite corregir una carencia como la que supone haber entrenado con puntos de tiempo únicos por sujeto y evaluar con varios puntos de tiempo de un mismo sujeto.

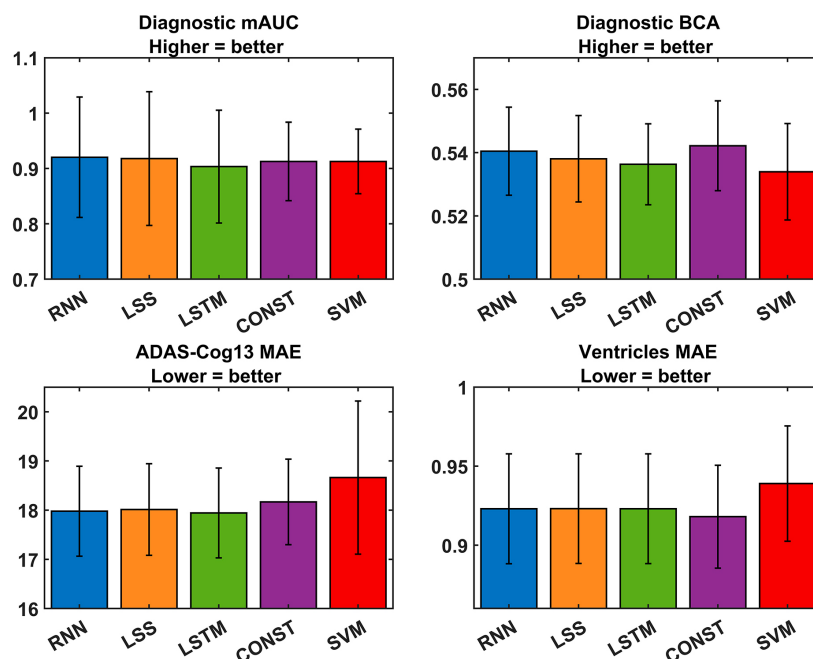


Figura 4.3: Rendimiento de los modelos de cada algoritmo promediado en 10 conjuntos de pruebas con la base de datos nueva y evaluado con la base de datos antigua. Las barras de error muestran el error estándar en los conjuntos de pruebas. Para el diagnóstico clínico, los valores más altos de mAUC y BCA indican un mejor rendimiento. Para ADAS-Cog13 y ventrículos, un MAE más bajo indica un mejor rendimiento. En aras de la brevedad, denotamos la RNN mínima como RNN, he indicamos que la SVM/SVR corresponde a 1 punto de tiempo.

Si continuamos observando la figura 4.3 nos damos cuenta que para la ADAS-Cog13 y el volumen ventricular empeora drásticamente la predicción frente a los resultados obtenidos en las figuras 4.1 y 4.2. Por ello ciertamente se puede afirmar que la métrica en la ADAS-Cog13 es directamente dependiente del número de puntos de tiempo de un sujeto, y que a mayor número de puntos por parte de un solo sujeto, mejor será la predicción, como hemos comentado en la subsección 4.2.2.

4.3. Resolución

En primer lugar, podemos destacar la importancia de tener una base de datos robusta y homogénea con una gran cantidad de sujetos en la que se puedan representar todos los diferentes grados de la enfermedad del Alzheimer y con una gran cantidad de números de puntos de tiempo por sujeto permite obtener resultados más globales de la capacidad de clasificación de los distintos modelos, además de demostrar que el empleo de mínima RNN como mejor algoritmo frente a otros como la SVM/SVR para la predicción longitudinal de biomarcadores multimodales de la enfermedad del Alzheimer y el diagnóstico clínico de los participantes hasta 6 años en el futuro, no es concluyente. Es por ello, que cabe destacar notablemente el buen desempeño llevado a cabo por la SVM/SVR con el empleo de un punto de tiempo junto con la mínima RNN empleando el rellenado de modelo, a pesar de que en el empleo de SVM/SVR es más propenso a que los datos se desperdicien porque los enfoques estáticos generalmente asumen que los participantes tienen el mismo número de puntos de tiempo de entrada, y en nuestro caso el problema de predicción longitudinal es más dinámico. Otro problema con los modelos estáticos es que la relación entre las características de entrada y los resultados podría variar con el tiempo, por lo que se podría lograr un mejor rendimiento mediante la construcción de modelos separados para predecir el mes 12, el mes 18, y así sucesivamente. En nuestro caso, construimos múltiples modelos SVM/SVR separados para predecir en un número fijo de puntos de tiempo futuros y realizamos la interpolación en puntos de tiempo intermedios. Por el contrario, los modelos basados en el estado, como pueden ser los modelos RNN mínima, LSS o LSTM, son más elegantes gracias a que manejan participantes con diferente número de puntos de tiempo y pueden, en principio, predecir un número ilimitado de puntos de tiempo en el futuro.

En segundo lugar, a pesar de no haber sido posible demostrar en los experimentos por obtener resultados muy similares, la variación y correcta elección de los hiperparámetros es una de las fases más determinantes para obtener resultados diferenciadores como sucede en el experimento original haciendo uso de la base de datos primigenia, donde destaca mínima RNN como mejor algoritmo para la predicción longitudinal.

En cambio, la actualización de la base de datos proporcionada por ADNI con datos nuevos, donde destacaban mayormente el incremento de nuevos sujetos con un único punto de entrada y con una mayor variedad de casos de estudio de los diferentes grados de la enfermedad del Alzheimer permite desarrollar una mejora en la predicción longitudinal tanto en el diagnóstico clínico como en el volumen

ventricular de los participantes hasta 6 años en el futuro, obteniendo el mejor resultado a fecha de 20 de junio de 2021 si participase en la tabla de clasificación de TADPOLE *Challenge* y se permitiese la actualización de la base de datos. En definitiva, la actualización de la base de datos del modo propuesto propiciaría a desarrollar algoritmos con una predicción más realista, permitiéndonos enfocar mejor el problema de fondo que es predecir mejor la evolución de los casos en los pasos intermedios en la enfermedad del Alzheimer. Como único inconveniente destaca el deterioro levemente en la predicción de la ADAS-Cog13.

Además, aunque el conjunto de datos proporcionado por ADNI incluía participantes con múltiples puntos de tiempo, para que el algoritmo sea clínicamente útil, debe tener éxito al tratar con la falta de datos y los participantes con un solo punto de tiempo de entrada. De este modo, observamos que el enfoque de usar el modelo para rellenar los datos que faltan, mediante la técnica del relleno del modelo, obtiene muy buenos resultados frente el uso de enfoques de preprocesamiento, como son el relleno hacia adelante o el relleno lineal. Por ende, el enfoque de llenado del modelo se puede considerar una forma de completar la matriz, ya que el RNN (o LSS) se entrena para minimizar la pérdida predictiva lo que equivale a maximizar la probabilidad de los datos de entrenamiento. Así pues, este método asume dependencias temporales entre las filas de la matriz de datos, donde cada fila es un punto de tiempo.

CONCLUSIONES Y TRABAJO FUTURO

5.1. Conclusiones

El Alzheimer es una de las enfermedades más extendidas en el siglo XXI y la prevención es fundamental para entenderla mejor y evitar que sea confundida con otras enfermedades neurodegenerativas. Esto ha originado que a día de hoy sea objeto de investigación por una gran número de organismos públicos y privados con el fin de detectar las causas que la producen, para así conocer como poder tratarla en un futuro.

El objetivo principal de este trabajo ha consistido en comprobar la utilización de las redes neuronales recurrentes, en particular el algoritmo de RNN mínima, para la mejora de la predicción en la detección de la enfermedad del Alzheimer mediante la ampliación de los datos proporcionados. De este modo, se ha demostrado a pesar de los problemas vinculados a la red y su adaptación con la nueva base de datos que este tipo de redes son una herramienta determinante para el estudio de la enfermedad del Alzheimer y una gran ayuda para los especialistas en la prevención y diagnóstico de la enfermedad. Además, se ha dejado de manifiesto que la ampliación de la base de datos permite desarrollar mejores soluciones y predecir mejor gracias a la utilización de un mayor número de datos que reflejen los grados intermedios de la enfermedad del Alzheimer permitiendo dar una visión más realista de la enfermedad que permita detectarla a tiempo.

Desde un primer momento, se han detectado problemas derivados de la correcta adaptación y selección de los algoritmos propuestos a estudio, frente a la modificación y variación de los datos proporcionados, que podrían empeorar la resolución definitiva de los algoritmos planteados. Además, se han abordado dificultades como la pérdida de datos de predicción, proponiendo distintas soluciones entre las que se encuentra el modelo RNN mínima y SVM/SVR que demuestran ser capaces de dar buenos resultados con datos longitudinales, frente al uso de otros algoritmos.

Como resultado de ello se han adaptado otra clase de algoritmos más comunes para proporcionar nuevos enfoques en su uso y permitir compararlos correctamente con el algoritmo central de este proyecto, junto con los problemas derivados de la falta de datos propiciados al tener un gran número de datos de diversos centros recopilados a lo largo de los años. De este modo, se ha demostrado

que tanto el modelo mínima RNN como el modelo SVM/SVR han funcionado razonablemente bien utilizando un punto de tiempo de entrada, lo que sugiere que el enfoque también podría funcionar para los datos transversales, y que la ampliación de la base de datos beneficia al mejor desempeño del modelo.

Además, se ha entendido el reto principal de la correcta detección de una enfermedad con un carácter cuantitativo como puede ser el caso de la medición de la proteína *Adas13*, pero también subjetivo, como puede suponer el uso del diagnóstico clínico previo para predecir la evolución de la enfermedad.

En definitiva, hemos concluido que la adaptación de nuevos datos junto con el uso del modelo RNN mínima y la SVM/SVR, permite una mejor predicción después del entrenamiento con datos longitudinales siendo más elegantes al manejar participantes con diferentes números de puntos de tiempo y pudiendo predecir un número ilimitado de puntos de tiempo en el futuro. Aunque para poder realizar una predicción eficaz de la demencia por Alzheimer es necesario detectarlo en una fase temprana del proceso de la enfermedad, debido a que hay mucho margen de mejora a la hora de predecir a largo plazo. En nuestro caso, en el mejor modelo propuesto la detección del *Adas13* en el año 6 decae considerablemente. Por lo tanto, se necesita una mejora significativa para la utilidad clínica.

5.2. Trabajo futuro

A la vista de los resultados obtenidos, podemos confirmar que el método de análisis es solvente y puede ser solución a parte del problema de la predicción de la enfermedad del Alzheimer. Así pues se plantean las siguientes proposiciones a futuro:

- Una posible dirección futura es investigar nuevas características o biomarcadores. Por ejemplo, las derivadas de la resonancia magnética de difusión o del etiquetado de espín arterial. Estudios previos clínicos han sugerido que diferentes patrones de atrofia (más allá del lóbulo temporal) podrían influir en el declive cognitivo al principio del proceso de la enfermedad, por lo que las características de atrofia consideradas en este estudio podrían no ser óptimas.
- Investigar los inhibidores de la GSK-3 [30] y diferentes tipos de agentes dirigidos a la proteína fosfo-tau. Para lograrlo, también es necesario desarrollar paralelamente nuevos biomarcadores fiables para la detección temprana de la enfermedad.
- Mejorar el algoritmo de SVM/SVR de aprendizaje no profundo, debido a que actualmente el mejor equipo hasta la fecha ha hecho uso de un algoritmo similar.
- Probar y mejorar el algoritmo de RNN mínima para el uso de datos transversales debido al buen uso en la predicción con un solo punto de tiempo en datos longitudinales.

GLOSARIO, ACRÓNIMOS Y DEFINICIONES

- **ADNI:** *Alzheimer's Disease Neuroimaging Initiative*
- **ADAS13:** *Alzheimer's Disease Assessment Scale 13-item Cognitive Subscale*
- **ANN:** *Artificial Neural Network*
- **A β :** *Amyloid Beta*
- **BCA:** *Balanced Classification Accuracy*
- **CAD:** *Computer-Aided Diagnosis*
- **CN:** *Cognitive Normal*
- **CNN:** *Convolutional Neural Networks*
- **CPA:** *Coverage Probability Accuracy*
- **CSF:** *Cerebral Spinal Fluid*
- **EA:** *Alzheimer*
- **EMCI:** *Deterioro Cognitivo Leve Temprano*
- **HC:** *Healthy Control*
- **HMM:** *Hidden Markov Models*
- **LMCI:** *Deterioro Cognitivo Leve Tardío*
- **LSTM:** *Long Short-Term Memory*
- **LSS:** *Linear State Space*
- **mAUC:** *Multiclass Area Under the receiver operating Curve*
- **MAE:** *Mean Absolute Error*
- **MCI:** *Deterioro Cognitivo Leve*
- **MRI:** *Magnetic Resonance Imaging*
- **NL:** *Cognitive Normal*
- **NN:** *Neural Networks*
- **PET:** *Positron Emission Tomography*

- **RNN:** *Recurrent Neural Networks*
- **SMC:** *Subjective Memory Complaint*
- **SVM:** *Support Vector Machines*
- **SVR:** *Support Vector Regression*
- **WES:** *Weighted Error Score*

BIBLIOGRAFÍA

- [1] M. Nguyen, N. Sun, D. C. Alexander, J. Feng, and B. T. T. Yeo. Modeling alzheimer's disease progression using deep recurrent neural networks. In *2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 1–4. IEEE, 2018.
- [2] Felix A. Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to Forget: Continual Prediction with LSTM. *Neural Computation*, 12(10):2451–2471, 10 2000.
- [3] Yarin Gal and Zoubin Ghahramani. A theoretically grounded application of dropout in recurrent neural networks, 2016.
- [4] Razvan V. Marinescu, Neil P. Oxtoby, Alexandra L. Young, Esther E. Bron, Arthur W. Toga, Michael W. Weiner, Frederik Barkhof, Nick C. Fox, Stefan Klein, Daniel C. Alexander, and the EuroPOND Consortium. TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease. *arXiv e-prints*, page arXiv:1805.03909, May 2018.
- [5] Tammaryn Lashley, Jonathan M Schott, Philip Weston, Christina E Murray, Henny Wellington, Ashvini Keshavan, Sandrine C Foti, Martha Foiani, Jamie Toombs, Jonathan D Rohrer, et al. Molecular biomarkers of alzheimer's disease: progress and prospects. *Disease Models & Mechanisms*, 11, 05 2018.
- [6] N. Sadeghi, N. L. Foster, A. Y. Wang, S. Minoshima, A. P. Lieberman, and T. Tasdizen. Automatic classification of alzheimer's disease vs. frontotemporal dementia: A spatial decision tree approach with fdg-pet. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 408–411, May 2008.
- [7] Shen Lu, Yong Xia, Weidong Cai, David Dagan Feng, and Michael Fulham. Cross-cohort dementia identification using transfer learning with fdg-pet imaging. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1550–1554, April 2018.
- [8] Hongming Li and Yong Fan. Early prediction of alzheimer's disease dementia based on baseline hippocampal mri and 1-year follow-up cognitive measures using deep recurrent neural networks. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 368–371, April 2019.
- [9] Ministerio de Sanidad C. y. B. S. *Plan integral de Alzheimer y otras demencias (2019-2023)*. Madrid: Ministerio de Sanidad, Consumo y Bienestar Social, 2019.
- [10] Robert DiPietro and Gregory D. Hager. *Chapter 21 - Deep Learning: RNNs and LSTM In Handbook of Medical Image Computing and Computer Assisted Intervention*. The Elsevier and MICCAI Society Book Series. Elsevier, January 2019.

- [11] Jee Bang, Salvatore Spina, and Bruce L Miller. Non-alzheimer's dementia 1: Frontotemporal dementia. *Lancet (London, England)*, 386(10004):1672, 2015.
- [12] José A Monge-Argilés, Victoria Monge-García, Ruth Gasparini-Berenguer, César García-Pérez, Laura Gabaldón-Torres, Juan Salas-Felipe, and Carlos Leiva-Santana. Frecuente alteración de la vía amiloide en la demencia con cuerpos de lewy prodrómica. *Rev. neurol.(Ed. impr.)*, pages 89–93, 2019.
- [13] P.L. Rodríguez García and D. Rodríguez García. Diagnosis of vascular cognitive impairment and its main categories. *Neurología (English Edition)*, 30(4):223–239, 2015.
- [14] Kyle Strimbu and Jorge A Tavel. What are biomarkers? *Current opinion in HIV and AIDS*, 5:463, 11 2010.
- [15] ADNI Dataset, <http://adni.loni.usc.edu/data-samples/access-data/>, 2021.
- [16] Răzvan V Marinescu, Neil P Oxtoby, Alexandra L Young, Esther E Bron, Arthur W Toga, Michael W Weiner, Frederik Barkhof, Nick C Fox, Polina Golland, Stefan Klein, et al. Tadpole challenge: Accurate alzheimer's disease prediction through crowdsourced forecasting of future data. In *International Workshop on PRedictive Intelligence In MEdicine*, pages 1–10. Springer, 2019.
- [17] Momoyo Ito, Kazuhito Sato, Ikuro Namura, and Minoru Fukumi. Extraction of brain regions for image diagnosis of alzheimer-type dementia based on atrophy progress speeds. In *IEEE Nuclear Science Symposium Medical Imaging Conference*, pages 3579–3580, Oct 2010.
- [18] Kanghan Oh, Young-Chul Chung, Ko Woon Kim, Woo-Sung Kim, and Il-Seok Oh. Classification and visualization of alzheimer's disease using volumetric convolutional neural network and transfer learning. *Scientific Reports*, 9(1):1–16, 2019.
- [19] Minmin Chen. Minimalrnn: Toward more interpretable and trainable recurrent neural networks. *arXiv preprint arXiv:1711.06788*, 2018.
- [20] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997.
- [21] Thomas Laurent and James H. von Brecht. A recurrent neural network without chaos. *The Computing Research Repository (CoRR)*, abs/1612.06212, 2016.
- [22] Davide Cannizzaro, Marina Zafiri, Daniele Jahier Pagliari, Edoardo Patti, Enrico Macii, Massimo Poncino, and Andrea Acquaviva. A comparison analysis of ble-based algorithms for localization in industrial environments. *Electronics*, 9:44, 12 2019.
- [23] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 01 2009.

-
- [24] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [25] Minh Nguyen, Tong He, Lijun An, Daniel C. Alexander, Jiashi Feng, and B.T. Thomas Yeo. Predicting alzheimer’s disease progression using deep recurrent neural networks. *NeuroImage*, 222:117203, 2020.
- [26] Solale Tabarestani, Maryamossadat Aghili, Mehdi Shojaie, Christian Freytes, Mercedes Cabrerizo, Armando Barreto, Naphtali Rishe, Rosie E. Curiel, David Loewenstein, Ranjan Duara, and Malek Adjouadi. Longitudinal prediction modeling of alzheimer disease using recurrent neural networks. In *2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pages 1–4, 2019.
- [27] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv e-prints*, page arXiv:1912.01703, December 2019.
- [28] Tanyaluck Thientunyakit, Chakmeedaj Sethanandha, Weerasak Muangpaisan, Orasa Chawalparit, Kuntarat Arunrungvichian, Tossaporn Siriprapa, Yudthaphon Vichianin, Swatabdi Kamal, Chaiyawat Suppasilp, Thonnapong Thongpraparn, Rujaporn Chanachai, and Juri Gelovani. Relationships between amyloid levels, glucose metabolism, morphologic changes in the brain and clinical status of patients with alzheimer’s disease. *Annals of Nuclear Medicine*, 34, 03 2020.
- [29] Ilija Ilievski, Taimoor Akhtar, Jiashi Feng, and Christine Shoemaker. Efficient hyperparameter optimization for deep learning algorithms using deterministic rbf surrogates. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.
- [30] Carmen Sayas and Jesús Ávila. Gsk-3 and tau: A key duet in alzheimer’s disease. *Cells*, 10:721, 03 2021.

APÉNDICES

FICHERO *features* Y UTILIDADES DE LA RED NEURONAL

Las siguientes secciones describen el fichero *features* y las configuraciones necesarias para hacer correr la red neuronal. Las referencias y motivo de empleo ha sido explicado en la sección 3.2.

A.1. Explicación desarrollada de las *features*

La selección principal de estas 23 *features* reside en las consignas recibidas por parte de TADPOLE Challenge [16]. También por elección personal se han decidido agregar ciertos test clínicos subjetivos que permiten tener una predicción más amplia, no centrada exclusivamente en resultados clínicos objetivos.

Principales pruebas cognitivas (excluyendo subtipos): - pruebas neuropsicológicas administradas por un experto clínico:

Etiqueta	Descripción	Tipo de Dato
CDRSB	Clinical Dementia Rating Sum of Boxes	Numérico
ADAS11	Escala cognitiva de 11 ítems-AD	Numérico
ADAS13	Escala cognitiva de 13 ítems-AD	Numérico
MMSE	Mini-Mental State Examination	Numérico
RAVLT_imnmediate	Test de aprendizaje verbal auditivo de Rey (puntuaciones de respuesta inmediata, aprendizaje, olvido y porcentaje de olvido)	Numérico
RAVLT_learning		
RAVLT_forgetting		
RAVLT_perc_forgetting		
Moca	Instrumento de cribado cognitivo para detectar el deterioro cognitivo leve	Numérico
Ecog		Numérico
FAQ	Functional Activities Questionnaire	Numérico

Tabla A.1: Tabla de pruebas cognitivas de las *features* usada por la red.

MRI ROIs (Freesurfer) - medidas de la integridad estructural del cerebro:

Etiqueta	Descripción	Tipo de Dato
Entorhinal	El volumen de la corteza entorrinal	Numérico
Fusiform	El volumen del giro fusiforme	Numérico
Hippocampus	Volumen del hipocampo	Numérico
ICV	Intra Cranial Volume	Numérico
MidTemp	El volumen de la circunvolución temporal media	Numérico
Ventricles	Volumen de los ventrículos	Numérico
WholeBrain	Volumen de Brain	Numérico

Tabla A.2: Tabla de medidas de la integridad estructural del cerebro de las *features* usada por la red.

Medias de FDG PET ROI, AV45 PET ROI, AV1451 PET ROI:

Etiqueta	Descripción	Tipo de Dato
FDG	miden el metabolismo celular, donde las células afectadas por la EA muestran un metabolismo reducido	Numérico
AV45	mide la carga de amiloide-beta en el cerebro, donde el amiloide-beta es una proteína que se pliega mal (es decir, su estructura 3D no se construye correctamente), lo que conduce a la EA	Numérico

Tabla A.3: Tabla de las medias FDG PET, AV45 PET y AV1451 PET de las *features* usada por la red.

Valores pertenecientes a los biomarcadores del líquido cefalorraquídeo (CSF):

Etiqueta	Descripción	Tipo de Dato
ABETA_UPENNBBIOMK9_04_19_17	niveles de amiloide en el CSF	Numérico
TAU_UPENNBBIOMK9_04_19_17	niveles tau en el CSF,	Numérico
PTAU_UPENNBBIOMK9_04_19_17	niveles en proteína tau en el CSF	Numérico

Tabla A.4: Tabla de Valores pertenecientes a los biomarcadores del líquido cefalorraquídeo (CSF) de las *features* usada por la red.

A.2. Ficheros modificados en el TADPOLE Challenge

A continuación, describimos la actualización de ficheros necesaria para realizar la nueva base de datos. Todas las modificaciones se han realizado respetando las normas establecidas por ADNI y

TADPOLE [16]. Para mejor entendimiento vamos a explicar los pasos de ejecución del código para poder efectuar la base de datos final “TADPOLE_D1_D2.csv” expresados por los organizadores del Desafío TADPOLE para generar los conjuntos de datos TADPOLE. Para ello requeriremos tener acceso a MATLAB y el lenguaje de programación Python.

- 1.– Colocar todas los ficheros ADNI necesarios juntos en una carpeta, junto con todos los *scripts TADPOLE*
- 2.– Correr en MATLAB el *script TADPOLE_D2.m*
- 3.– Correr en python el *script TADPOLE_D1_D2.py*
- 4.– Correr en MATLAB el *script TADPOLE_D3.m*

Archivos Antiguos	Archivos Nuevos
ADNIMERGE	ADNIMERGE
ADNIMERGE_DICT	ADNIMERGE_DICT
ARM	ARM
DXSUM_PDXCONV_ADNIAL	DXSUM_PDXCONV_ADNIAL
UCSFFSL_02_01_16	UCSFFSL_02_01_16
UCSFFSL_DICT_11_01_13	UCSFFSL_DICT_11_01_13
UCSFFSL51ALL_08_01_16	UCSFFSL51ALL_08_01_16
UCSFFSL51ALL_DICT_05_04_16	UCSFFSL51ALL_DICT_05_04_16
UCSFFSX_11_02_15	UCSFFSX_11_02_15
UCSFFSX_DICT_08_01_14	UCSFFSX_DICT_08_01_14
UCSFFSX51_08_01_16	UCSFFSX51_11_08_19
UCSFFSX51_DICT_08_01_14	UCSFFSX51_DICT_02_04_14
BAIPETNMRC_09_12_16	BAIPETNMRC_04_12_18
BAIPETNMRC_DICT_09_12_16	BAIPETNMRC_DICT_11_14_17
UCBERKELEYAV45_10_17_16	UCBERKELEYAV45_05_12_20
UCBERKELEYAV45_DICT_06_15_16	UCBERKELEYAV45_DICT_06_15_16
UCBERKELEYAV1451_10_17_16	UCBERKELEYAV1451_10_17_16
UCBERKELEYAV1451_DICT_10_17_16	UCBERKELEYAV1451_DICT_10_17_16
DTIROI_04_30_14	DTIROI_04_30_14
DTIROI_DICT_04_30_14	DTIROI_DICT_04_30_14
UPENNBIOMK9_04_19_17	UPENNBIOMK9_04_19_17
UPENNBIOMK9_DICT_04_19_17	UPENNBIOMK10_DICT_07_29_19

Tabla A.5: Tabla de los ficheros modificados a la hora de realizar la actualización de la base de datos en TADPOLE Challenge. Los ficheros en **negrita** representan los ficheros actualizados de manera más visual para facilitar la lectura.

A.3. Columnas más importantes en TADPOLE Challenge

Para la simplificación del número de columnas destacables dentro del fichero *dataset* obtenido a la hora de realizar la ampliación y actualización de los valores del TADPOLE Challenge explicado en la sección 3.2.2 vamos a destacar los campos más importantes del fichero. Esto es debido a que obtenemos un fichero de 1907 columnas donde la gran mayoría carecen de importancia, están obsoletas ó sean de interés en un futuro.

Atributos que indican el ID del participante

- RID: ID del participante
- PTID: ID del participante

Atributos que indican el momento del examen del participante.

- EXAMDATE: fecha de examen
- EXAMDATE_bl: fecha del primer examen
- M: meses de examen, con una aproximación de 6 meses como continuo
- Month_bl: valor fraccionario de los datos de examen
- VISCODE: código de visita del participante, que indica los meses de examen

Atributos que indican el estado del participante

- DX: indica el estado de la enfermedad en el participante en 3 categorías distintas correspondientes a normales cognitivamente, deterioro cognitivo leve y demencia
- DX_bl: indica el estado de la enfermedad en el participante tras el primer examen

